







```

567 TCGTGGCTTCACCAACTATACGGATTGAGAGCTACACCTACTTC.. 614
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
149 CysCysGlyTTrpValSerPheIyrAsnIrrPhrAspAsnIaGluLeuE 165
615 ...AAGAGAAACAGTGCCTTCCCACTCTGT.....TGCA 648
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
165 fAsnIrrProGluValIrrThrProCysSerCysGluValIysGlyGlu 182
649 ATGACAAGCTTACCAACACAGCAGCAATGAACCTGC.....683
:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
182 IuAspAsnSerLeuSerValIrrGlySGlyPheCysGluIaProGlyAsn 198
684 ...ACCAAGCAAAAGGCTCAGCAGCAAAAA.....GTAGAGG 718
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
199 ArgThrIrrGlnSerGlyAsnIrrProGluAspTrpProValTyrGlnGlu 215
719 TTGCTTCACATACGCTTTGTGTGACATCCGAACATATGCACTACCCGTG 768
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
215 YCysMetIuIrrSVaIrrGlnIaIrrPheGlnIaGluAsnLeuGlyIleI 232
769 GTGGGTGTGGACAGCTGAATTGGGGGCTCGAGCTGGCTGCATATTGTG 818
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
232 euGlyValGlyValGlyValaIleIrrGluLeuLeuGlyMetValLeu 248
819 TCCATGTATCTGTACTGCAATCTACCAATAACTCCACTTTCGCTTGCCA 868
|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
249 SerIle.....CysLeuCysAr 254
869 CTAC 872
|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
254 gHis 255

```

```

seq_documentation_block:
C33/R2/IA4 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 16-Feb-1997
C:Accession: I49561
R:Nagira, M.; Imai, T.; Ishikawa, I.; Uwabe, K.I.; Yoshie, O.
Cell: Immunol. 157, 144-157, 1994
A:Title: Mouse homologue of C33 antigen (CD82), a member of the transmembrane 4 superfamily
A:Reference number: I49561; MUID:9431678
A:Accession: I49561
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-266 <RES>
A:Cross-references: GB:D14883; NID:g984036; PID:g498368
A:Superfamily: CD9 antigen

```

```
alignment_scores:
  Quality: 284.00
  Ratio: 1.753
  Percent Similarity: 58.273
  Length: 278
  Gaps: 10
  Percent Identity: 28.058
```

```
alignment_block:
US-09-030-606-111 x I49561 ..
```

Align seg 1/1 to: I49561 from: 1 to: 266

```

129 TGGTTAGCCTTCATTAAAGCACTGATGATCCCTTCATTGGCGACTT 178
    |||  :::  |||  :::  |||  :::  |||  :::  |||  :::  |||  :::
    5 CysValIysAlaThrIysTyrPheLeuPheLeuPheAsnLeuLeuPhePh  21
179 TCAGTGTGGTGCAGCCCTTGGCAAGTGGCATCGGGTGCATTCGATG  228
    :::  :::  :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
    21 eIleLeuGlyAlaValIleLeuGlyPheClyAlaTrpIleuAlaAspL  38
229 GGGCATCTCTTTCGGAAGACTCTCGGGCCACTCGTCCATCGGCATCGAG  278
    |||  :::  :::  :::  :::  :::  :::  :::  :::  :::  :::
    38 ysAsnSerPheIleSerValIleLeuGlnThrSerSerSerSer.....  51

```

279 TTGTGCAAGTGGGC...TACTTCCTCATCGCAGCGCGCGTGTGCTCT 325  
52 ...LenglnValglValatryValPheIleGlValglValleIleThrII 67  
326 TGCTCTGTGTTCTCTGGCGCTGATGTGCTTAAGACTGAGACCAAGTGTG 375  
67 eValMetgllyPheLeuGllycSIIleGlValalValaIasngIuValaIArgCylS 84  
376 CCTCGTGCAGCTTCTTCTTCATCCTCCTCCTCATCTTCAATGCTGAGGT 425  
84 euleuGllyLeuTyrPheValPheLeuLeuIleuLeuIleIaGlVal 100  
426 GCAGCTGCTGTGGCTGGCCTTGGCTGCTACACACATGCGTGACACTCTCT 475  
101 ThrValGllyAl... 105  
476 GAGCTGTGCTGTAGTGCCTGCATCAAGAAGATATAGTCCAG... 521  
105 uPheTyrPheAsnAlaAspLysLeuLysgluMetglYasnThrValM 122  
522 .....GAGACTTCACT.....CAA 536  
122 eAspIleIleArgAsnTyrThrAlaAsnAlaThrSerSerArgIuGlIn 138  
537 GTGTGGAAACCCACCATGAAAGCGCTCAAGTGTGCTGACC... 581  
139 AlatrPspTyrValglAlaIaGlValIyScysGlyTyrPValSerH 155  
582 ....AACTATACGGATTTTGAGAGACTCACCTACTTCAAGAAGAACGTG 627  
155 sTyrAsnTrpThrGluaSngIuIuMetglYpHeThrLys...Thrt 171  
628 CCTTGTCCCCATTCCTGT..... 644  
171 hTyrTyrProCysSerCysgluLysIleLysgluIuAspAsngIuLeuIle 187  
645 TGCATATGACACAGCTCACCAACACAGCCATGAAGCTGCACCAAGCAAA 694  
188 ValIlybLysglYpHeCysgluAlaAspAsnSerThrValSerGluaSns 204  
695 GGCTACAGACCAAAAAGTA.....GAGGCTTGCTTCAATCAAGCTTTGT 738  
204 nProgluaSptPrProValasnThrIuGlIyCysMetglYlalaGlua 221  
739 ATGACATCCGAACACTATGACAGTCACCGTGAGGAGTGTGACCTGGAAT 788  
221 latrPheugIuSngIuaSnpHeGlYIleuLeuLeuIyValIySalaGlYal 237  
789 GGGGGCCCTGCAGCTGGCTGCATGATTTGTCCATGTCTGTACTGCAA 838  
238 AlavallIleGlIuLeuLeuGlYleuPheLeuSerIle..... 249  
839 TCTACATAAAGTCCACTTCTGCTCTGCCACATAC 872  
550 .....CysLeuSyrArgTyr 254

```

seq_name: p1r1:139368
seq_documentation_block:
T-cell acute lymphoblastic leukemia associated antigen 1 - human
N:Alternate names: cell surface glycoprotein (clone A15); TALIA-1
C:Species: Homo sapiens (man)
C:Date: 23-Feb-1996 #sequence_revision 23-Aug-1996 #text_change 13-Sep-1997
C:Accession: I39368; EMBL:154784
R:Emil, N.; Kitehorst, K.; Seto, M.; Ueda, R.; Saito, H.; Takahashi, T.
Immunogenetics 37, 193-198, 1993
A:Title: Isolation of a novel cDNA clone showing marked similarity to ME491/CD63 super
A:Reference number: I39368; MIMD:93131291
A:Accession: I39368
A:Molecule type: mRNA
A:Residues: 1-244 <res>
A:CROSS-references: GB:D10653; NID:G285900; PID:G285901
A:Experimental source: Immature T cell line HPB-ALL

```

R:Takegi, S.; Fujikawa, K.; Imai, T.; Fukuhara, N.; Fukudome, K.; Minegishi, M.; Tsuchi  
Int. J. Cancer 61, 706-715, 1995  
A:Title: Identification of a highly specific surface marker of T-cell acute lymphoblasti  
A:Reference number: 154784; MUID:95286314  
A:Accession: 154784  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-244 <RE2>  
A:Cross-references: GB:D29808; NID:9475005; PID:9475006  
C:Genetics:  
A:Gene: GDB:MXS1; DXS1692E; A15; TALLA-1  
A:Cross-References: GDB:202921  
A:Map position: Xq11-Xq11  
C:Superfamily: CD9 antigen  
C:Keywords: glycoprotein; surface antigen; transmembrane protein  
F:1-1/Domain: intracellular #status predicted <CY1>  
F:12-35/Domain: transmembrane #status predicted <TM1>  
F:36-51/Domain: extracellular #status predicted <EX1>  
F:52-76/Domain: transmembrane #status predicted <TM2>  
F:77-80/Domain: intracellular #status predicted <CY2>  
F:81-102/Domain: transmembrane #status predicted <TM3>  
F:103-207/Domain: extracellular #status predicted <EX2>  
F:208-233/Domain: transmembrane #status predicted <TM4>  
F:233-244/Domain: intracellular #status predicted <CY3>  
F:49,150,153,172,183/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment\_scores:  
Quality: 280.50 Length: 245  
Ratio: 1.775 Gaps: 6  
Percent Similarity: 64.490 Percent Identity: 26.939

alignment\_block:  
US-09-030-606-111 x 139368 ..

Align seg 1/1 to: 139368 from: 1 to: 244

```

120 ACCATGACGAGTCTGACCTTCATTAAGACCATGATGATCCTTCATTT 169
    |||          |||||          |||||          |||||
3   Thrlysprovallierhrcysleuylthrlleuulelleltyrserph 19
170 GCTCATCTTTCTGTGTGTCAGCCCTGTTGGCAGTGGGCTGGGTGT 219
    |||||          |||||          |||||          |||||
19   evalphtripilerhrylvalilleleuualvalaiglyaltrpilyl 36
220 CAATCGATGGGGCATCTTCTTGAAGATCTTGGGGCCACGTGCTCCAGT 269
    |||||          |||||          |||||          |||||
36   ysleuthrlleuylthrylleserleuilealguasnserthr... 51
270 GCCATGCACTTTGTCACAGTGGGCTACTCTCTCATCCAGCGGCGGTGT 319
    |||||          |||||          |||||          |||||
52   .....AsnAlaProtyrValleulellellythrlythrth 63
320 GGTCTTGTCTGTTGTTCTGCGGCTGCTATGTCATGATGAGTGAAGCA 369
    |||||          |||||          |||||          |||||
63   rillevalvalpheelglyleupheglycyspheelalthrCysarglyserp 80
370 AGTCTGCCCTCGTACGCTTCTCTCATCCCTCCCTCATCTTATGTCT 419
    |||||          |||||          |||||          |||||
80   rotrpmetleuylsleuyltralaletpheelserleuvalpheeluala 96
420 GAGTTGACAGTCTGTCGTGTCGCTTGGTGTACACCAATGCTGAGCA 469
    |||||          |||||          |||||          |||||
97   Gluleuvalalae glylleser glypheelalphearghsguilelysas 113
113   pthpheeluarthrlytrhraspalamec...Glnthryrasnelya 129
470 CTTCCTGACGCTGCTGTAGTCTGCCATCAAGAAAGATTATGCTCC 519
    |||||          |||||          |||||          |||||
520 AGGAAGCTTCACATCAAGTGTGGAACACACCAATGAAGGCTCAAGTGC 569
129 snaspGluArgSerArGAlaValaspHisValGlnArgSerLeuSerCys 145

```

570 TGTGCTTCCACCACTATACGATTTGAGAGCTCACCTACTTCAAGA 619  
||||| ||||| ||||| |||||  
146 CysglyValaGlnasnrythrhrasntpserThrserProtyrPheuleu 162  
620 GAACAGTCCCTTCCCATCTTCTGTCATGACCAAGTCACACACAG 669  
||||| ||||| ||||| |||||  
162 uhts...glylleProProserCys.....M 171  
670 CCATGAAGCC...TGCACCAAGCAAGCTCAGCACCAGCAAGTA... 713  
||||| ||||| ||||| |||||  
171 etasnGluThrAspCysAsnProGlnAspleuHisasnleuthrValala 187  
714 .....GAGGTTGCTTCATCAGCTTTTGTATGACAT 745  
||||| ||||| ||||| |||||  
188 AlaThrlyValasnGlnlysglytyrAspleuValthrserpeme 204  
746 CCGACATATGACGATCCGTGGTGTGTCAGCTGGAATGGGGCC 795  
||||| ||||| ||||| |||||  
204 tGluThrAsnmetGlyllelealaglyValaIapheglyllealaphes 221  
796 TCGAGCTGGCTGCCATGATTTGTCCATGATCTG 830  
||||| ||||| ||||| |||||  
221 erGlnleuileglyMetleuValaIacysCysleu 232  
seq\_name: p1r1:A36056

seq\_documentation\_block:

tumor-associated antigen CO-029 - human  
C:Species: Homo sapiens (man)  
C:Date: 16-Nov-1990 #sequence-revision 09-Aug-1996 #text-change 05-Sep-1997  
C:Accession: A36056  
R:Szala, S.; Kasal, Y.; Steplewski, Z.; Rodeck, U.; Koprowski, H.; Linnenbach, A.J.  
Proc. Natl. Acad. Sci. U.S.A. 87, 6833-6837, 1990  
A:Title: Molecular cloning of cDNA for the human tumor-associated antigen CO-029 and  
A:Reference number: A36056; MUID:90370878  
A:Accession: A36056  
A:Molecule type: mRNA  
A:Residues: 1-237 <SZ>  
A:Cross-references: GB:M35252; NID:g180925; PID:g180926  
C:Superfamily: CD9 antigen  
C:Keywords: glycoprotein; transmembrane protein  
F:2-11/Domain: intracellular #status predicted <CY1>  
F:12-33/Domain: transmembrane #status predicted <TM1>  
F:34-52/Domain: extracellular #status predicted <EX1>  
F:53-78/Domain: transmembrane #status predicted <CY2>  
F:79-82/Domain: intracellular #status predicted <CY3>  
F:83-107/Domain: transmembrane #status predicted <TM3>  
F:108-205/Domain: extracellular #status predicted <EX2>  
F:206-233/Domain: transmembrane #status predicted <TM4>  
F:233-237/Domain: intracellular #status predicted <CY3>  
F:37,118/Binding site: carbohydrate (asn) (covalent) #status predicted

alignment\_scores:  
Quality: 277.50 Length: 238  
Ratio: 1.790 Gaps: 7  
Percent Similarity: 65.126 Percent Identity: 30.252

alignment\_block:  
US-09-030-606-111 x A36056 ..

Align seg 1/1 to: A36056 from: 1 to: 237

```

141 ATTAAGACCTGATGATCCCTTCATTTGCTCATCTTTCTGTGTGTC 190
    |||||          |||||          |||||          |||||
8   IlelystyriserMetPheThrPheasnPheleuphrtrpLeuysglyl 24
191 AGCCCTGTGTCAGTGGGATCGGATGTCGATGATGAGTGGGATCCTTTC 240
    |||||          |||||          |||||          |||||
24   eleulleuAlaIeulalalletrpAlarGyalserAsnAspSer...G 40
241 TGAAGATCTTGGGCGACGTGCTCCAGTGCATGCACTTGTGCAACGTG 290
    |||||          |||||          |||||          |||||

```





Melanoma-associated antigen CD63 - human  
 A:Alternate names: antigen ME491; Lysosomal membrane glycoprotein CD63; ME491/CD63 antigen  
 C:Species: Homo sapiens (man)  
 C:Date: 17-May-1996 #sequence, revision 09-Aug-1996 #text-change 26-Feb-1999  
 C:Accession: J36016; S01418; A39514; B35826; A61173; A56782  
 R:Hotta, H.; Miyamoto, H.; Hara, I.; Takahashi, N.; Homma, M.  
 Biochem. Biophys. Res. Commun. 185, 436-442, 1992  
 A:Title: Genomic structure of the ME491/CD63 antigen gene and functional analysis of the  
 A:Reference number: J36016; MUID:92287132  
 A:Accession: J36016  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-238 <RES>  
 A:Cross-references: EMBL:X62654; NID:9430755; PID:9430756  
 R:Hotta, H.; Ross, A.H.; Huebner, K.; Isobe, M.; Wendeborn, S.; Chao, M.V.; Ricciardi, R.  
 Cancer Res. 48, 2955-2962, 1988  
 A:Title: Molecular cloning and characterization of an antigen associated with early stag  
 A:Reference number: S01418; MUID:88210273  
 A:Accession: S01418  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <HOT>  
 A:Cross-references: EMBL:X07982; NID:934526; PID:934527  
 R:Merzlekar, M.J.; Wijngaard, P.L.J.; Peters, P.J.; Slama, J.J.; Nieuwenhuis, H.K.; Cley  
 J. Biol. Chem. 266, 3239-3245, 1991  
 A:Title: CD63 antigen. A novel lysosomal membrane glycoprotein, cloned by a screening pr  
 A:Reference number: A39514; MUID:91131632  
 A:Accession: A39514  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <MET>  
 A:Cross-references: GB:M58485  
 R:Rapp, G.; Freudenstein, J.; Klaudiny, J.; Mucha, J.; Wempe, F.; Zimmer, M.; Schell, K.  
 DNA Cell Biol. 9, 479-485, 1990  
 A:Title: Characterization of three abundant mRNAs from human ovarian granulosa cells.  
 A:Reference number: A35826; MUID:91025550  
 A:Accession: B35826  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <RAP>  
 A:Cross-references: GB:M59907; NID:9189383; PID:9189384  
 A:Note: the authors did not translate the codons for residues 205 through 224  
 R:Azorsa, D.O.; Hymn, J.A.; Hildreth, J.E.K.  
 Blood 78, 280-284, 1991  
 A:Title: CD63/PLtp40: a platelet activation antigen identical to the stage-specific, me  
 A:Reference number: A61177; MUID:91300080  
 A:Accession: A61177  
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 2-68, 'P', '70-238 <AZO>  
 R:Hildreth, J.E.K.; Derr, D.; Azorsa, D.O.  
 Blood 77, 121-132, 1991  
 A:Title: Characterization of a novel self-associating Mr 40,000 platelet glycoprotein.  
 A:Reference number: A61173  
 A:Accession: A61173  
 A:Molecule type: protein  
 A:Residues: 2-8, 'X', '10-16, 'XX', '19-21 <HIL>  
 R:Wang, M.X.; Earley Jr., J.J.; Shields, J.A.; Donoso, L.A.  
 Arch. Ophthalmol. 110, 399-404, 1992  
 A:Title: An ocular melanoma-associated antigen. Molecular characterization.  
 A:Reference number: A56782; MUID:92181348  
 A:Accession: A56782  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <WAN>  
 A:Cross-references: GB:S93788; NID:9246538; PID:9246539  
 A:Experimental source: uveal melanoma  
 A:Note: sequence extracted from NCBI backbone (NCBI:93788, NCBI:P.93790)  
 C:Genetics:  
 A:Gene: GDB:CD63; MIAL  
 A:Cross-references: GDB:I20186; OMIM:155740  
 A:Map position: 12q12-12q13  
 A:Introns: 22/3; 85/3; 110/3; 142/3; 189/3; 217/3  
 C:Superfamily: CD9 antigen  
 C:Keywords: glycoprotein; lysosome; surface antigen; transmembrane protein  
 F:2-38/Product: melanoma-associated antigen ME491 #status experimental <MAT>  
 F:2-11/Domains: intracellular #status predicted <CYL>

Alignment_scores:	Quality:	249.00	Length:	269
Ratio:	1.556		Gaps:	12
Percent Similarity:	59.480	Percent Identity:	28.253	
alignment_block:				
US-09-030-606-111 x I38016 ..				
Align seg 1/1 to: I38016 from: 1 to: 238				
123	ATGACGCTCTTACGCTTCATTAGACCATGATGATCTCTTCATTTGCT	172		
7	MetScysVallysphe.....	LeuLeuTyValleu	18	
173	CATCTTTCTGTGTGTGCAGCC.....	CTGTGGCAGTGGCATCTGGG	216	
18	UleuLaIaPheCysAlaCysAlaValAlaGlyLeuIleAlaValAlaGlyVal.....	33		
217	TGTCAATGATGGGGCAGTCCTTTCTGGAAGATCTTCGGGGCAGTCGTCTCC	266		
34	.....GlyAlaLeuIleVal.....	LeuSerGln	41	
267	ACTGCCATGACGTTGTTCACGTGGGCTACTCTCTC.....	AT	304	
42	ThrIleIleGlnGlyAlaIaThrProGlySerLeuLeuProValValIleIle	58		
305	GCCACCGCGGCTGTGTGCTTGTGCTTGGTTCCGTGGGCTGCTATGGTG	354		
58	eAlaValAlaGlyValPheLeuPheLeuValAlaPheValAlaGlyCysGlyValA	75		
355	CTAAGACTGAGAGCAAGTGTCCCTCTGTGACGTTCTTCTTCACCTCTCTC	404		
75	IaCysGlyGlnAsnTyrcysLeuMetIleIlePheAlaIlePheLeuSer	91		
405	CTTCATCTTCATTTGCTGAGGTTTGACGCTGTGTGTGCTGCCCTTGGTATC..	452		
92	LeuIleMetLeuValAlaValAlaAlaIleLeuIleGlyTyraPheAr	108		
453	....ACCCAAATGGCTGAG.....	CACTTCCTGACGTGCTGGTAG	489	
108	gaSpLySValMetSerGlnPheAsnAsnAsnPheArgGlnGlnMet....	123		
490	TGCCTGCATCAAGAAAGATTATGTTCCAGAGAACTTCACCTCAAGTG	539		
124	.....GlnAsnTyrcProLySAsnAsnIleThrAlaSerIle	135		
540	TGGAACACACCATGAAGGCTCAAGTGTGTGTGCTTCCACCACTATAC	569		
136	LeuAspArgMetGlnAlaAspPheIlyCysCysGlyAlaAlaAsnTyrrH	152		
590	GGATTGTGAGACTCACCCCTACTTCAAGAGAAACAGAGCTTTCGCCCAT	639		
152	rsrPrpGluIlyIleProSerMetSerIlySAsnArgVal...ProAspS	168		
640	TCGTGTGAATGACCAAGCTCACCAACACGCAATGAACCTGCACCAAG	689		
168	ercYscys...IleAsnValThrValAlaGlyScGlyIleAsnPheAsnGlu	183		
690	CAAAAGGCTCAAGCAAAAGTGAAGGAGGTTCCTCAATAGCTTTTGTGA	739		
184	LySAlaIleHsIlyS.....	GlnGlyCysValGluIlySLeuGly	197	

```

740 TCACATCCGAACATAATGACAGTCACCCCTGGGTGGTGTGGCAGCTGGAAATG 789
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
197 YTRPLeuAglYsAsnValLeuValAlaAlaAlaAlaLeuGlyIleA 214
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
790 GGGCGCTCGAGCTGGCTGCCATGATGTGCCATGTATCTGATCTCAAT 839
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 LRPheValGluValLeuGlyIleValPheAla..... 224
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
840 CTACATTAAGTCCACTTCTGCTGCGACCTACTCTGCCACATGGGAAAC 889
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 .....CysCysLeuValYsSe 230
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
890 TGTGAAG 896
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
230 rIleArg 232
      ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

seq_name: pI1:A39574

seq_documentation_block:
L: leukocyte antigen OX-44 - rat
C: Species: Rattus norvegicus (Norway rat)
C: Date: 30-Aug-1991 #sequence_revision 09-Aug-1996 #text_change 05-Sep-1997
C: Accession: A39574
R: Bellacosa, A.; Lazo, P.A.; Bear, S.E.; Tschlis, P.N.
Mol. Cell. Biol. 11, 2864-2872, 1991.
A: Title: The rat leukocyte antigen MRC OX-44 is a member of a new family of cell surface
A: Reference number: A39574; MUID:91203909
A: Accession: A39574
A: Molecule type: mRNA
A: Residues: 1-219 <BRL>
A: Cross-references: GB:M57276; NID:g205897; PID:g205898
C: Superfamily: CD9 antigen
C: Keywords: glycoprotein; transmembrane protein
F: 1-10/Domain: intracellular #status predicted <CY1>
F: 11-36/Domain: transmembrane #status predicted <TM1>
F: 37-54/Domain: extracellular #status predicted <EX1>
F: 55-73/Domain: transmembrane #status predicted <TM2>
F: 74-79/Domain: intracellular #status predicted <CY2>
F: 80-106/Domain: transmembrane #status predicted <TM3>
F: 107-181/Domain: extracellular #status predicted <EX2>
F: 182-204/Domain: transmembrane #status predicted <TM4>
F: 205-219/Domain: intracellular #status predicted <CY3>
F: 119,129,148/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment_scores:
      Quality: 242.00      Length: 243
      Ratio: 1.603      Gaps: 6
      Percent Similarity: 62.140      Percent Identity: 24.691

Alignment_block:
US-09-030-606-111 x A39574 ..

Align seg 1/1 to: A39574 from: 1 to: 219

123 ATGCAAGTCCTCAGCTTCATTAGACCATGATGATCCTTCATTTGCT 172
      ||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
3 MetSerSerLeuLysLeuLeuLysTyrValLeuPhePheAsnPheLeu 19
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
173 CATCTTCTGTGTGTGTCAGCCCTGTGGCAGTGGGCATCTGGGTCAA 222
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
19 uRPheTrpValCysGlyCysAspIleLeuGlyPheGlyIleAsnLeuVal 36
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
223 TCGATGGGCGATCCTTTCGAAAGATCTTGGCGCATCTGTCGTCACATGCC 272
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
36 aRGlnAsnThr.....TyrGlyIleLeuPheArgAsn... 46
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
273 ATGCAAGTTGTCAACGTGGGCTACTTCTCATTCGACGCGCGGCTGTGTGT 322
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
47 LeuProPheLeuThrLeuGlyAsnValLeuValIleValGlySerIleI 63
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
323 CTTTGTCTTGCTTCTGGCTGCGCTCATGTGCTGAAGATGAGAGCAACT 372
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

63 ekeetValValAlAAlpheLeuglYcysMecelYserIlelyGsluaSnlySC 80

373 GTGCCCTCGTGAAGTTCTTCATCCCTCCTCATCTTCATTCTGTGAG 422  
||| :||||| :  
80 ysleuMetsSerPhePhValleuleuLeuileuleuMlaSlu 96  
423 GTTGCAAGCTGTGGTGCCTTGGTGTACACCACAATGGCTGACACTT 472  
||| :||||| :  
97 ValThrLeuAlAlleuleuPheValTyR.....Glulysly 109  
473 CCGAGCTGTGGTGTGCTGCTGCCATCAAG.....AAAGATTyG 513  
:||||| :  
109 sIleAsnThrLeuValAlaGlulYleuAsnMsperIlegInHsTyH 126  
514 GTTCCAGAGAAGACTTCACCTCAAGTGTGGAMACCCACCATGAAGGGCTC 563  
||| :||||| :  
126 IsSerTsPrnsInserrThrArgMetAlaTrpAsPheIleInsErGlnleu 142  
564 AAGTGTGTGGCTTACACCAATAACGATTTTGAGAGCTACCCCTACTT 613  
:||||| :  
143 GlnCySgCyglYvalAsnglySerAsrPyrlIsesrgly..... 156  
614 CAAGAAGAACAGTGCCTTTCGCCCATTTCTGTGTGCATAGCAACGTCACCA 663  
157 .....ProPseSerGysPsoSeergYala.... 165  
664 ACACAGCCAATGAMAACCTGCACCAAGCAAAGAGCTCACGCCAAAAGTA 713  
166 .....Aspyal 167  
714 GAGGTTGCTTCATCAGCTTTGTATGACATCCGAACTAATGCACTAC 763  
:||||| :  
168 GlnGlyCyStylYslYsglyGlnAlaTrpPheHisSerAsnPheluty 184  
764 CGNGTGGTGTGGCAGCTGGAATGGGGGCGCTCGAGCTGGTGCATGA 813  
184 rIleGlylleValThrIleCysValCysValIleGlnValleuGlyMets 201  
814 TTGTGTGCATGTATCTGTACTGCATCTA 842  
201 erPheAlaleuthrleuAsnCysGlnIle 210

seq\_name: p1r1:A43522

seq\_documentation\_block:  
23K integral membrane protein - fluke (Schistosoma mansoni)  
N.Alternate names: Sm23  
C.Species: Schistosoma mansoni  
C.Date: 28-Oct-1992 #sequence\_revision 09-Aug-1996 #text\_change 05-Sep-1997  
C.Accession: A43522  
R.Wright, M.D.; Henkle, K.J.; Mitchell, G.F.  
J. Immunol. 144, 3195-3200, 1990  
A.Title: An immunogenic M-r 23,000 integral membrane protein of Schistosoma mansoni v  
A.Reference number: A43522; MUID:90217533  
A.Accession: A43522  
A.Molecule type: mRNA  
A.Residues: 1-218 <MRI>  
A.Cross-references: GB:M4453; NID:g161029; PID:g161030  
A.Note: Observations of molecular weight suggest this protein is at most lightly glyco  
C.Keywords: glycoprotein; transmembrane protein  
F.1-13/Domain: intracellular #status predicted <CY1>  
F.14-36/Domain: transmembrane #status predicted <TM1>  
F.37-56/Domain: extracellular #status predicted <EX1>  
F.57-77/Domain: transmembrane #status predicted <TM2>  
F.78-82/Domain: intracellular #status predicted <CI2>  
F.83-108/Domain: transmembrane #status predicted <TM3>  
F.109-183/Domain: extracellular #status predicted <EX3>  
F.184-206/Domain: transmembrane #status predicted <TM4>  
F.207-218/Domain: intracellular #status predicted <CY3>  
F.165/Binding site: carbohydrate (Asn) (covalent) #status predicted







Accession: JX0221  
 A:Molecule type: mRNA  
 A:Residues: 1-226 <MAR>  
 A:Cross-references: GB:M81720; NID:9162820; PID:9162821  
 A:Experimental source: ocular ciliary epithelial cell  
 C:Superfamily: CD9 antigen  
 C:Keywords: glycoprotein; transmembrane protein  
 F:2-226/Product: CD9 antigen #status predicted <MAR>  
 F:2-11/Domain: intracellular #status predicted <CY1>  
 F:36-53/Domain: transmembrane #status predicted <TM1>  
 F:36-53/Domain: extracellular #status predicted <EX1>  
 F:54-76/Domain: transmembrane #status predicted <TM2>  
 F:77-80/Domain: intracellular #status predicted <CY2>  
 F:81-109/Domain: transmembrane #status predicted <TM3>  
 F:110-192/Domain: extracellular #status predicted <EX2>  
 F:193-219/Domain: transmembrane #status predicted <TM4>  
 F:220-226/Domain: intracellular #status predicted <CY3>  
 F:50/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment\_scores:  
 Quality: 235.00 Length: 258  
 Ratio: 1.577 Gaps: 11  
 Percent Similarity: 57.752 Percent Identity: 28.682

Alignment\_block:  
 US-09-030-606-111 x JX0221 ..

Align seg 1/1 to: JX0221 from: 1 to: 226

```

105 CCCTGAAGAGAGCCACATGACGCTTCGCTTCATTAAACCATGAT 154
    ||| ||||| ||| ||||| ||||| ||||| ||||| |||||
2 Provalylglylythr..LysCys.....Ileuylrleu 14
155 GATCCTTCATTTTCCTCATCTTTGTGTGTGTCGACGCCCTGTGGCAG 204
    ::: |||||:|||||:|||||:|||||:|||||:|||||:|||||
14 upheglypheasnPheliepherpleuValaglyIlealavalLeuSery 31
205 TGGGCGATGGGTCATCATGATGGGGCATCTTCTGAAATCTTCGGG 254
31 aGlyleuTrpleuArpHeasPserGlnThr...LysSerIlePheGlu 46
255 CCAGTGTCTCCAGTCCATGACGATTGTCAAGTGGCTACTTCTCAT 304
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
47 GlnGlnAsnAsnAspSerSerPheTrpGlnGlyVal...TyrIleu 62
305 CGCAGCCGGCGTTGTGTCTTGTCTTGTGTTCCTGGGCTGCTATG 354
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
62 eGlyalaglyAlaLeuMetMetLeuValGlyPheLeuGlyCysGlyA 79
355 CTAAAGTCAGAGCAAGTGGCCCTGTGACGTTCTTCTTCAATCCCT 404
    ||| ||||| ||||| ||||| ||||| ||||| |||||
79 lavalalnIuSerGlnCysMetLeuGlyLeuPhePheSerPheLeu 95
405 CTGATCTTCATGCTGAGGTTGCAGCTGCTGTGCTGCTGTGCTG 454
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 ValIlePheAlaIleGlyValAlaAlaAlaIleTrpGly.....Tyr 110
455 CACAAATGGCTGAGCACTTCTGACGTTGCTGAGTACCTGTCATCA 504
    : ||||| ||||| ||||| ||||| ||||| |||||
110 rHslsYsgluGlu.....ValIleYsgluValGln 121
505 AAGATATATGTTCCCAAGAACTTCACATCAAGTGTGAAC..... 545
    || ||||| :|||||:|||||:|||||:|||||:|||||
121 ySerHetr.....GluAspThrTyrAsnYsLeuYsAsnYsAspGlu 135
546 .....ACCACCATGAAG.....GGGCTCAAGTGTGTG 574
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
136 ProGlnArgGluThrLeuYsAlaIleHisIleAlaLeuAspCysG 152
575 CTTCACCAACTATAGGATTTTGAAGAGTCACTCAATTCAAAGACA 624
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 yLeuThrGlyValPProGlu..... 158

```

```

625 GTGCCCTTCCCATTCATCTGTTCATATGACAAAGTCACCAACCAAT 674
159 .....GlnPheLeuThr 162
675 GAACCTGCACCAAGCAAAAGGCTCAGCACCA.....AAGTAGAGGTT 721
    :|||||:|||||:|||||:|||||:|||||:|||||
163 AsPThrCysProProLysAsnLeuIleAspSerLeuYsThrArgProcy 179
722 CTTCATATGACCTTTGTATGACATCCGAACATATACAGTACCGGCG 771
    : |||||:|||||:|||||:|||||:|||||:|||||
179 sProGlnAlaIleAspGluIlePheArgSerLysPheHisIleIleGlyA 196
772 GTGTGCACCTGGAATTTGGGGCTCGAGCTGCTGCATGATTTGTGCC 821
    :|||||:|||||:|||||:|||||:|||||:|||||
196 lValGlyIleIleValIleAlaValIleMetIlePheGlyMetValPheSer 212
822 ATGTATCTGTACTGCAATCTCAAA 845
    ||| ||| ||| |||||
213 MetIleLeuCysAlaIleArg 220

```

seq\_name: p1r1:A46508

seq\_documentation\_block:

CD63/ME491 antigen homolog - rat  
 N:Alternate names: mast cell antigen AD1  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 18-Jun-1993 #sequence, revision 09-Aug-1996 #text, change 05-Sep-1997  
 C:Accession: A46508; S16776  
 R:Nishikata, H.; Oliver, C.; Mergenhagen, S.E.; Straganian, R.P.  
 J. Immunol. 149, 862-870, 1992  
 A:Title: The rat mast cell antigen AD1 (homologue to human CD63 or melanoma antigen M  
 A:Reference number: A46508; MUID:92340890

A:Accession: A46508  
 A:Molecule type: mRNA  
 A:Residues: 1-238 <NIS>  
 A:Cross-references: EMBL:X61654; NID:955601; PID:955602  
 A:Note: this antigen was found localized to mast cells in tissue samples, but was ind  
 A:Note: sequence extracted from NCBI backbone (NCBI:109346, NCBI:109349)  
 C:Comment: This heavily glycosylated protein of 50-60K (27K after deglycosylation) is  
 C:Superfamily: CD9 antigen  
 C:Keywords: glycoprotein; lysosome; mast cell; surface antigen; transmembrane protein  
 F:1-11/Domain: intracellular #status predicted <CY1>  
 F:12-35/Domain: transmembrane #status predicted <TM1>  
 F:36-51/Domain: extracellular #status predicted <EX1>  
 F:52-76/Domain: transmembrane #status predicted <TM2>  
 F:77-80/Domain: intracellular #status predicted <CY2>  
 F:81-103/Domain: transmembrane #status predicted <TM3>  
 F:104-202/Domain: extracellular #status predicted <EX2>  
 F:203-228/Domain: transmembrane #status predicted <TM4>  
 F:228-238/Domain: intracellular #status predicted <CY3>  
 F:130,150,172/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment\_scores:  
 Quality: 235.00 Length: 265  
 Ratio: 1.546 Gaps: 10  
 Percent Similarity: 57.358 Percent Identity: 26.792

Alignment\_block:  
 US-09-030-606-111 x A46508 ..

Align seg 1/1 to: A46508 from: 1 to: 238

```

123 ATGCAAGTCTTCAGCTTCATTAAACATGATGATCTTCAATTTGCT 172
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
7 MetLysCysValLysPhe.....LeuLeuTyrValIle 18
173 CATCTTTCGTGTGTGTCAGCC.....CTGTGTCAGTGGGCACTGGG 216
    :|||:|||||:|||||:|||||:|||||:|||||:|||||
18 uLeuAlaPheCysAlaCysAlaValGlyLeuIleAlaIleGlyValAla 35
217 TGTCAATGATGGGCAATCTTCTGAAAGATCTTGGGCACTGTCGTC 266

```



```

191 .....HISGLNLYSLSPQLULEPNHESERGLYLSLEUTRYLE 204
      ||
743 CATCCGAACATAATGCAGTCCACCGGTGGTGTGCGAGCTGGAGATTGGGG 792
      ||| |||||::|||::: ||:
204 uTLEGLYLEALAAVLEVALVALEVALLEMET..... 216
      |||
793 GCGCTGACCTGGCTGCCATGATTGTGTGCATGTATCTACTGCATCTA 842
      :::: |||||:::||||| ||| |||:::
217 .....ILEPHEGLUETILELEUSEMELVALEULENCYSCGLYLE 230
      ::::
843 CAA 845
      ::
231 Arg 231

```

This Page Blank (uspto)

OM of: US-09-030-606-111 to: SwissProt\_37.\* out\_format: pfs  
Date: Sep 25, 1999 4:45 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+np2.model -DEV=xlp  
-O=/cgrn2.1/uspro.spool/us09030606/runeat\_24091999\_171618\_29883/app-query.fasta.1  
-DS=SwissProt\_37 -QRT=fastan -SUFFIX=rsb -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELPOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blonsum62  
-TRANS=human40.ccl -LIST=45 -DOCALIGN=200 -THR.SCORE=esscore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:  
Query: US-09-030-606-111  
Query length: 1289  
Database: SwissProt\_37.\*  
Database sequences: 77977  
Database length: 28268293  
Search time (sec): 155.430000

score\_list:  
Sequence Strd Orig ZScore EScore Len Documentation  
SwissProt\_37:CD37\_MOUSE + 305.00 444.90 6.8e-18 281 | 061470 mus musculus (mouse)  
SwissProt\_37:NG2\_HUMAN + 301.00 444.34 1.4e-17 238 | 014817 homo sapiens (human)  
SwissProt\_37:CD37\_RAT + 296.00 435.41 3.8e-17 281 | P31053 rattus norvegicus (rat)  
SwissProt\_37:CD82\_HUMAN + 294.50 433.61 5.1e-17 267 | P27701 homo sapiens (human)  
SwissProt\_37:CD82\_MOUSE + 284.00 417.91 3.8e-16 266 | P41732 mus musculus (mouse)  
SwissProt\_37:A15\_HUMAN + 280.50 413.41 7.4e-16 244 | P41705 homo sapiens (human)  
SwissProt\_37:CD02\_HUMAN + 277.50 409.17 1.3e-15 237 | P19075 homo sapiens (human)  
SwissProt\_37:A15\_MOUSE + 273.50 402.92 2.9e-15 244 | O62283 mus musculus (mouse)  
SwissProt\_37:CD37\_HUMAN + 271.00 397.95 4.7e-15 281 | P11049 homo sapiens (human)  
SwissProt\_37:CD53\_HUMAN + 269.00 397.12 6.7e-15 219 | P19357 homo sapiens (human)  
SwissProt\_37:CD53\_MOUSE + 264.50 389.12 1.6e-14 233 | O35566 mus musculus (mouse)  
SwissProt\_37:CD53\_HUMAN + 263.00 388.17 2.1e-14 218 | 061451 mus musculus (mouse)  
SwissProt\_37:CD63\_HUMAN + 249.00 366.46 3.2e-13 237 | P88962 homo sapiens (human)  
SwissProt\_37:CD51\_HUMAN + 244.50 359.15 7.6e-13 253 | P48509 homo sapiens (human)  
SwissProt\_37:CD53\_RAT + 242.00 356.70 1.2e-12 218 | P2485 rattus norvegicus (rat)  
SwissProt\_37:IM23\_SCHNA + 237.50 349.96 2.9e-12 218 | P19331 schistosoma mansoni  
SwissProt\_37:CD81\_HUMAN + 237.50 349.27 2.9e-12 236 | P18582 homo sapiens (human)  
SwissProt\_37:CD63\_RABIT + 236.00 346.98 3.8e-12 237 | O28709 oryctolagus cuniculus  
SwissProt\_37:CD9\_BOVIN + 235.00 345.94 4.6e-12 225 | P30932 bos taurus (bovine)  
SwissProt\_37:CD63\_RAT + 235.00 345.49 4.7e-12 237 | P28648 rattus norvegicus (rat)  
SwissProt\_37:CD81\_MOUSE + 232.00 341.03 8.3e-12 236 | P35762 mus musculus (mouse)  
SwissProt\_37:CD9\_RAT + 228.00 335.45 1.8e-11 225 | P40241 rattus norvegicus (rat)  
SwissProt\_37:CD63\_MOUSE + 228.00 335.00 1.8e-11 237 | P41731 mus musculus (mouse)  
SwissProt\_37:CD9\_MOUSE + 223.00 327.96 4.6e-11 225 | P40240 mus musculus (mouse)  
SwissProt\_37:IM23\_SCHNA + 221.00 325.23 6.8e-11 218 | P27591 schistosoma japonicum  
SwissProt\_37:CD9\_HUMAN + 217.00 318.89 1.5e-10 227 | P21926 homo sapiens (human)  
SwissProt\_37:IM23\_SCHNA + 215.50 316.99 2.0e-10 218 | O26499 schistosoma haematobium  
SwissProt\_37:CD9\_FELCA + 215.00 315.97 2.2e-10 227 | P30409 cercopithecus aethiops  
SwissProt\_37:CD9\_CEREA + 215.00 315.89 2.2e-10 227 | P30409 cercopithecus aethiops  
SwissProt\_37:UPKA\_BOVIN + 178.50 260.09 2.4e-07 258 | P38572 homo sapiens (bovine)  
SwissProt\_37:UPKA\_HUMAN + 169.50 246.66 1.4e-06 258 | O00332 homo sapiens (human)  
SwissProt\_37:UPKB\_BOVIN + 151.00 218.85 4.8e-05 229 | P38573 bos taurus (bovine)  
SwissProt\_37:UPKB\_MOUSE + 139.50 201.62 0.0004 229 | P30443 mus musculus (bovine)  
SwissProt\_37:YKRB\_CAEEL + 111.00 159.43 0.1044 244 | P24285 caenorhabditis elegans  
SwissProt\_37:YL11\_CAEEL + 109.50 155.13 0.1431 309 | O11098 caenorhabditis elegans  
SwissProt\_37:RSL1\_XENLA + 109.00 153.43 0.1595 345 | O42581 xenopus laevis (afric  
SwissProt\_37:RDS\_CHICK + 106.50 149.46 0.2586 354 | O42281 gallus gallus (chicken)  
SwissProt\_37:FIG2\_YEAST + 103.50 131.82 0.5461 1609 | P25653 saccharomyces cerevisiae  
SwissProt\_37:YVZ\_XANCV + 101.00 134.32 0.8141 784 | P14788 xanthomonas campestris  
SwissProt\_37:KML\_HUMAN + 99.50 136.91 1.02 449 | P15544 homo sapiens (human)  
SwissProt\_37:KML\_MOUSE + 99.50 136.91 1.02 449 | P25561 mus musculus (mouse)  
SwissProt\_37:PSA7\_DICDI + 98.50 145.65 1.08 138 | P11022 dictyostellium discoideum

SwissProt\_37:RDS2\_CHICK + 99.00 137.98 1.10 364 | O42282 gallus gallus (chicken)  
SwissProt\_37:KIR4\_HUMAN + 99.00 135.17 1.14 503 | P36897 homo sapiens (human)  
SwissProt\_37:RDS2\_XENLA + 97.50 136.20 1.45 345 | O42582 xenopus laevis (afric)

seq\_name: SwissProt\_37:CD37\_MOUSE

seq\_documentation\_block:  
ID CD37\_MOUSE STANDARD; PRT; 281 AA.

AC 061470:  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE LEUCOCYTE ANTIGEN CD37.  
GN CD37.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC RODENTIA; SCIROGAMATHI; MURIDAE; MORINAE; MUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SV; TISSUE=LIVER;  
RX MEDLINE: 97002026  
RA TOLINSON M.G., WRIGHT M.D.,  
RT "Characterisation of mouse CD37: cDNA and genomic cloning."  
RL MOL. IMMUNOL. 33:867-872(1996).  
CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.  
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE  
ANTIGENS (TM4 SUPERFAMILY).

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
-----  
DR EMBL: U18372; G755242; JOINED.  
DR EMBL: U18367; G755242; JOINED.  
DR EMBL: U18368; G755242; JOINED.  
DR EMBL: U18369; G755242; JOINED.  
DR EMBL: U18370; G755242; JOINED.  
DR EMBL: U18371; G755242; JOINED.  
DR MGI: 88330; CD37.  
DR PROSITE: PS00421; TM4; 1.  
DR PFAM: PF00335; transmembrane4; 1.  
KW GLYCOPROTEIN; ANTIGEN; TRANSMEMBRANE; SIGNAL-ANCHOR.  
FT DOMAIN 1 12  
FT TRANSMEM 13 36  
FT DOMAIN 37 59  
FT DOMAIN 60 74  
FT TRANSMEM 75 85  
FT DOMAIN 86 111  
FT TRANSMEM 112 241  
FT DOMAIN 242 266  
FT TRANSMEM 267 281  
FT CARBOHYD 170 170  
FT CARBOHYD 183 183  
FT CARBOHYD 188 188  
SQ SEQUENCE 281 AA; 31852 MW; EDAE4017 CRC32;

alignment\_scores:  
Quality: 305.00 Length: 286  
Ratio: 1.784 Gaps: 7  
Percent similarity: 59.790 Percent identity: 28.671

alignment\_block:  
US-09-030-606-111 x CD37\_MOUSE ..

Align seg 1/1 to: CD37\_MOUSE from: 1 to: 281  
117 GCACCATCAGTGCCTTCATTAGACCATGATCCTTCATCA 166  
||||| :|||:|||||:||||| :|||:|||||:|||||

```

3  AlagInGluSerCysLeuSerLeuIleLysTyrPheLeuPheValPheAs 19
167 TTGCGATCATCTGTGTGTGTCAGCCCTGTGGCAGTGGGATCTGGG 216
19 nleuPhePhePheValLeuGlyGlyLeuIlePheCysPheIleThrTrpI 36
217 TGTCAATGATGGGCGATCTTCTGGAAGATCTTCGGGCGCATCTGCTCC 266
36 leuIleAspLysThrSerPheValSerPheValGly..... 48
267 AGTGCATGCAAGTTGTCAACGTG.....GGCTACTTCTTCATCGC 307
49 .....LeuSerPheValProLeuGlnThrTrpSerLysValLeuAlaVa 63
308 AGCGGGGGTGTGCTTGTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTA 357
63 IserGlyValLeuThrMetAlaLeuAlaLeuLeuGlyCysValIleValAl 80
358 AGACTGAGACAGAGTGCCTCGTGCAGCTTCTTCATCTCTCTCTCTC 407
80 euLysGluLeuArgCysLeuLeuGlyLeuLysPheGlyMetLeuLeu 96
408 ATCTTCATTGCTGAGGTGCAAGCTGCTGTGCTGCTGCTGCTGCTGCTAC 457
97 leuPheAlaThrGlnIleThrLeuGlyIleLeuIleSerThrGlnArgVa 113
458 AATGCTGAGACACTTCTGACGTTGCTGCTGCTGCTGCTGCTGCTCAAGAAG 507
113 IArgLeuGlnArgValGlnGlnLeuValLeuArgThrIle...Glns 129
508 ATTATGTTCCCGAGAGACTTCACT.....CAAGTGTGAGACACC 548
129 eryTyrArgThrAsnProAspLysIleThrAlaGlnGluProPheValPtyr 145
549 ACCATGAAAGGCTCAAGTGTGTGCTGCTTCCCACTATACGATTTTGA 598
146 AlAlaGlnPheGlnLeuArgCysGlyTyrPheIleSerProArgAspTyrAs 162
599 GAGTCACTCCACTTCAAGAGAGACAGTGCCTTCCCGCATTC..... 641
162 nLysAlaGlnMetLeuLysAlaAsnGluSerGluProPheValPro 179
642 .....TGTTCATGACACGCTACCAAC..... 665
179 ySerCysTyrAsnSerThrAlaThrAsnAspSerThrValPheAspLys 195
665 ..... 665
196 leuPhePheSerGlnLeuSerArgLeuGlyProArgAlaLysLeuArgL 212
666 .ACAGCCATGAAACCTGACACCAAGCAAAAGGCTCAGACCAAGCAAGTAG 714
212 nThrAlaAspIleCysAlaLeuProAlaLysAlaHisIleTyrArg...G 228
715 AGGGTCTTCATCAGCTTTGTATGATGATGATGATGATGATGATGATGATG 764
228 IArgCysAlaGlnSerLeuGlnLysTrpLeuHisAsnAsnIleIleSer 244
765 GTGGGTGTGTGGCAGCTGCAATGGGGGCTGCAAGTGGCTGCTGCTGCTG 814
245 ILeuValGlyIleCysLeuGlyValGlyLeuLeuGlyLeuGlyPheMetL 261
815 TGTGTGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 864
261 rLeuSerIlePheLeuCysArgAsnLeuAspHisValTyrAspArgLeuA 278
865 GCCACTAC 872
278 IArgTyr 280
seq_name: SwissProt_37:NAG2_HUMAN
seq_documentation_block:

```

```

ID NAG2_HUMAN STANDARD: PRT: 238 AA.
AC 014817:
DT 15-JUL-1998 (REL. 36, CREATED)
DT 15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NOVEL ANTIGEN 2 (NAG-2).
GN TM4SF7 OR NAG2.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA, METAZOA, CHORDATA, VERTEBRATA, MAMMALIA, EUTHERIA.
OC PRIMATES, CATARRHINI, HOMINIDAE, HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 98030601.
RA TACHIBANA I., BOBOROVA J., BERDICHEVSKI F., ZUTTER M.M., HEWLER M.E.;
RT "NAG-2, a novel transmembrane-4 superfamily (TM4SF) protein that
RT complexes with integrins and other TM4SF proteins."
RL J. BIOL. CHEM. 272:29181-29189(1997).
CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.
CC -1- SUBUNIT: FORMS A COMPLEX WITH INTEGRINS.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN MULTIPLE TISSUES BUT IS ABSENT IN
CC BRAIN, LYMPHOID CELLS, AND PLATELETS.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE
CC ANTIGENS (TM4 SUPERFAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF022813; G2586350; -.
DR MIM: 602644; -.
DR PROSITE: PS00421; TM4; 1.
DR PFAM: PF00335; transmembrane4; 1.
KW GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL-ANCHOR.
KW DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT (POTENTIAL).
FT DOMAIN 35 55 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 56 76 POTENTIAL.
FT DOMAIN 77 85 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 86 106 POTENTIAL.
FT DOMAIN 107 201 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 202 222 POTENTIAL.
FT DOMAIN 223 238 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 152 152 POTENTIAL.
FT CARBOHYD 161 161 POTENTIAL.
FT SEQUENCE 238 AA; 26118 MW; E0D8AF65 CRC32;

alignment_scores:
Quality: 301.00 Length: 247
Ratio: 1.942 Gaps: 9
Percent Similarity: 62.753 Percent Identity: 29.555

alignment_block:
US-09-030-606-111 x NAG2_HUMAN ..
Align seg 1/1 to: NAG2_HUMAN from: 1 to: 238

129 TGTGTGATGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 178
|||||:.....:|||||:.....:|||||:.....:
5 CysLeuGlnAlaValLysTyrLeuMetPheAlaPheAsnLeuLeuPheTr 21
179 TGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 228
|||||:.....:|||||:.....:|||||:.....:
21 pleuGlyGlyCysGlyValLeuGlyValGlyIleTrpLeuAlaAlaThrG 38
229 GGGCATCTTCTGAAAGATCTTGGGCGCATCTTCAGTCCAGTCCAGTCCAG 278
:::|||||:.....:|||||:.....:
38 lngIySerPheAlaThrLeu.....SerSerSerPhePro 49

```

```

CC      ANTIGENS (TM4 SUPERFAMILY).
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X53517; G55912; -.
DR      PIR; B47629; B47629.
DR      PIR; JCI501; JCI501.
DR      PROSITE; PS00421; TM4; 1.
DR      PRAM; PF00335; Transmembrane4; 1.
CC      GLYCOPROTEIN; ANTIGEN; TRANSMEMBRANE; SIGNAL-ANCHOR.
FT      DOMAIN 1 12
FT      TRANSSEM 13 36
FT      DOMAIN 37 59
FT      TRANSSEM 60 74
FT      DOMAIN 75 85
FT      TRANSSEM 86 111
FT      DOMAIN 112 241
FT      TRANSSEM 242 265
FT      DOMAIN 267 281
FT      CARBOHYD 170 170
FT      CARBOHYD 183 183
FT      CARBOHYD 188 188
FO      SEQUENCE 281 AA; 31750 MW; 31DAD084 CRC32;
CC      -----
alignment_scores:
      Quality: 296.00      Length: 286
      Ratio: 1.741      Gaps: 7
      Percent Similarity: 59.441      Percent Identity: 28.322
alignment_block:
US-09-030-606-111 x CD37_RAT ..
Align seg 1/1 to: CD37_RAT from: 1 to: 281
117 GCCACCATGCAAGTCTTCAGCTTCATTAAGACCATGATCTCTTCA 166
    ||| ::::::::::::::::::::|
    3 AaagIngIuSerCySleuSerLeuIleIlyStyrPheLeuPhenValaIphas 19
167 TTTCATCATCTTCTGTGTGGTGAGACCCCTGTGGAGAGGCGCATCTGG 216
    |||::::::::::::::::::|
    19 nleuPhePhePheValleuGlyGlyleuIlePheCysPheGlyThrTyrI 36
217 TGTCAATCATGATGGGCAATCTTTCGAAGATCTTGCGGCGACATGTCTCC 266
    ::| ||||| ::::::::::::::::::::|
    36 leuIleIleSplysrHsrPheValSerPheValGly..... 48
267 AGTGCATGACACTTGTCAACGTG.....GGCTACTTCTCATGCC 307
    ::::::::::||| |::| ::| ::| ::|
    49 .....LeuSerPheValaProleuGlnThrTrpSerLysValleuSerVa 63
308 AGCGGCGGCTGTGGCTTGTGCTTGGTTCCTGGGCGCTGATGGCTGA 357
    ::::::::::::::::::::|
    63 lSerIlyalIleuThrMetAlaLeuAlaLeuIleuGlyCysValGlyAla 80
358 AGACTGAGAGCAAGTGTGCCCTGTGACAGCTTCTTCATCTCTCTCTC 407
    ||| ::||| ||| ::||| ::|||
    80 euIyGgluIleuArGcysleuIleuGlyLeuTyrPheGlyMetleuIleu 96
408 ATCTTCATGTCTGAGGCTGACAGCTGTGTGGTGGCTTGCTGATACAC 457
    ::||| ::||| ::||| ::||| ::|||
    97 leuPheAlaThrGlnIlePheIleuGlyIleuIleSerThrGlnArGy 113
458 AATGGCTGAGCACTTCTGACAGCTGTGTGATGCTGATGCTGCATCAGA 507
    : ||::: ::| ::||| ::||| ::|||
    113 lArGleuGluArGArGValGlnIleuValleuArGThrIle...Glns 129

```





```

21 eileleuglyalaVal11leleuglypheglyVal1trp1leuAlaaspl 38
229 GGGGACCTTTCAGAGATCTTGGGACGACGATCGCCAGCGCATGAG 278
   :::::::::::::::::::::::::::: ||| |||
38 ysSerSerPheIleSerValleuInhrSerSerSer..... 51
279 TTGTCAACGTGGC...TACTTCATCATCGACCGCGCTGTGTGCTT 325
   ::::::::::||| ||| ::::::::::::::::::::|||
52 ...LeuArgMetGlyAlaTyrValPheIleGlyValGlyAlaValTlme 67
326 TGCCTTGGTTCCTGGCTGCTTGGTGCCTAACAGTGAAGCAAGTGTG 375
   ::::::::::::::::::::||| ||| ::::::
67 tleuMetGlyPheleuGlyCysIleGlyAlaValAlaGlnValArgCysL 84
376 CCGCTGAGACCTTCTTCATCCTCCGTCATCTTCATCTGAGAGT 425
   ||| :::::: ::::::::::::::::::::|||
84 euleuGlyLeuTyrPheAlaPheleuLeuLeuIleleuIleAlaGlnVal 100
426 GCAGCTCTGTGTGCTGCTGTGTGACACCAATGGCTGACATTCCT 475
   ::::::::::||| ||| ::::::::::::::::::::
101 ThrAlaGly.....AlaLeuPheTyrPheAsnMetGlyLysLeuLysG 115
476 GACCTTG...CTGTAGTGCCTGCATCAGAAAGATTAT...GCTTCC 519
   ::::::::::::::::::::||| ::::::::::::::::::::
115 ngluMetGlyGlyLeuValThrGluLeuIleArgAspTyrAsnSerSer 132
520 AGGAACCTTCATCA...GTGTGAACACCAACCATGAAGGCTCAG 566
   ::::::::::||| ||| ::::::::::::::::::::
132 rgluAspSerleuGlnsPalaTrpAspTyrValGlnAlaGlnValLys 148
567 TGCTGTGCTTACCACTATACGAGATTGAGAGCTACCCCTACTTC.. 614
   ||| ||| ::::::::::::::::::::|||
149 CysGlyGlyTyrPalaSerPheTyrAsnTrpThrAspAsnAlaGluLeu 165
615 ...AAGAGACAGTGCCTTCCCCCATTCGT.....TGCA 648
   ::::::::::::::::::::|||
165 tAsnArgProGluValThrTyrProCysSerCysGluValLysGlyGlu 182
649 ATGCAACGTTCACACACAGCCCAATGAACCTC..... 683
   ||| ||| ::::::::::::::::::::
182 luAspAsnSerleuSerValArgLysGlyPheCysGluAlaPheGlyAsn 198
684 ...ACCAAGCAAAAGCTCAGACCAAA.....GTAGAGG 718
   ||| ::::::::::::::::::::|||
199 ArgThrGlnSerGlyAsnHisPheGluAspTrpProValTyrGlnGlu 215
719 TTGCTTAATCAGCTTTGTATGACATCCGAACTAAATGCACTACCGT 768
   ||| ::::::::::::::::::::|||
215 yCyMetGlyLysValGlnAlaTrpLeuGlnGluAsnLeuGlyIleIle 232
769 GTGTGTGGGACGCTGAATGGGGCCCTCGAGCTGCGCTCATATGTG 818
   ||| ::::::::::::::::::::|||
232 euglyValGlyValGlyValAlaIleIleGlyLeuLeuGlyMetValLeu 248
819 TCCATGATCTGTACTGCAATCTCAATAAGTCCACTTCTGCTCTGCA 868
   ||| ::::::::::::::::::::|||
249 SerIle.....CysLeuCysArg 254
869 CTAC 872
254 GH15 255
seq_name: SwissProt_37:CD82_MOUSE
seq_documentation_block:
ID CD82_MOUSE STANDARD; PRT; 266 AA.
AC P40237;
DT 01-FEB-1995 (REL. 31, CREATED)
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE CD82 ANTIGEN (INDICIBLE MEMBRANE PROTEIN R2) (C33 ANTIGEN) (1A4).
GN CD82.
OS MUS MUSCULUS (MOUSE).

```

```

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA MEDLINE; 94313678.
RX NAGIRA M., IMAI T., ISHIKAWA I., UMABE K.I., YOSHIE O.:
RT "Mouse homologue of C33 antigen (CD82), a member of the transmembrane
RT 4 superfamily: complementary DNA, genomic structure, and
RT expression."
RL CELL. IMMUNOL. 157:144-157(1994).
CC -1- FUNCTION: ASSOCIATES WITH CD4 OR CD8 AND DELIVERS COSTIMULATORY
CC SIGNALS FOR THE TCR/CD3 PATHWAY.
CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE SPLEEN AND THE
CC KIDNEY. LOW EXPRESSION IN SKELETAL MUSCLE AND IN THE HEART.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE
CC ANTIGENS (TM4 SUPERFAMILY).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
CC EMBL: D14883; G498368; -.
CC MGD: MG1:104651; CD82.
CC DR PROSITE; PS00421; TM4; 1.
CC DR PFM; PFM0335; transmembrane4; 1.
CC KW GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL-ANCHOR; ANTIGEN.
CC FT DOMAIN 1 10 10
CC FT TRANSMEM 11 34
CC FT DOMAIN 35 53
CC FT TRANSMEM 54 72
CC FT DOMAIN 73 83
CC FT TRANSMEM 84 110
CC FT DOMAIN 111 227
CC FT TRANSMEM 228 249
CC FT DOMAIN 250 266
CC FT CARBOHYD 127 127
CC FT CARBOHYD 131 131
CC FT CARBOHYD 157 157
CC FT CARBOHYD 197 197
CC SQ SEQUENCE 266 AA; 29628 MW; AFE2BBEA CRC32;
-----
alignment_scores:
Quality: 284.00 Length: 278
Ratio: 1.753 Gaps: 10
Percent Similarity: 58.273 Percent Identity: 28.058
-----
alignment_block:
US-09-030-606-111 x CD82_MOUSE ..
Align seg 1/1 to: CD82_MOUSE from: 1 to: 266
-----
129 TGCTTCAGCTTCAATTAGACCATGATGATCTCTCAATTGCTCATCTT 178
   ||| ::::::::::||| ::::::::::::::::::::|||
5 CysValLysValThrLysTyrPheLeuPheLeuPheAsnLeuPhePh 21
21 eileleuglyalaVal11leleuglypheglyVal1trp1leuAlaaspl 38
179 TCTGTGTGTGACAGCCCTGTGTGACAGTGGCATCTGGGTTCATCATG 228
   ::::::::::::::::::::||| ||| ::::::::::::::::::::
229 GGGACCTTTCAGAGATCTTGGGACGACGATCGCCAGCGCATGAG 278
   ||| ||| ::::::::::::::::::::|||
38 ysAsnSerPheIleSerValleuInhrSerSerSer..... 51
279 TTGTCAACGTGGC...TACTTCATCATCGACCGCGCTGTGTGCTT 325
   ::::::::::||| ||| ::::::::::::::::::::|||
52 ...LeuArgMetGlyAlaTyrValPheIleGlyValGlyAlaValTlme 67

```

[illegible]

	seq_name: SwissProt_37:A15_HUMAN	seq_documentation_block:	
ID	A15_HUMAN	STANDARD;	PRT; 244 AA.
AC	P41732;		
DT	01-NOV-1995 (REL. 32, CREATED)		
DT	01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	CELL SURFACE GLYCOPROTEIN A15 (T-CELL ACUTE LYMPHOBLASTIC LEUKEMIA ASSOCIATED ANTIGEN 1) (TALLA-1) (MEMBRANE COMPONENT, X CHROMOSOME)		
DE	ASSOCIATED ANTIGEN 1) (TALLA-1) (MEMBRANE COMPONENT, X CHROMOSOME)		
DE	SURFACE MARKER 1).		
GN	TM4SF2 OR MXS1 OR A15.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA.		
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.		
RN	1 1		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=PERIPHERAL BLOOD LYMPHOCYTES;		
RA	MEDLINE; 93131291.		
EM	EMI N. KITAOKI K., SETO M., UEDA R., SAITO H., TAKAHASHI T.,		

RT "Isolation of a novel cDNA clone showing marked similarity to  
RT ME491/CD63 superfamily."  
RT  
RL IMMUNOGENETICS 37:193-198(1993).  
RN [2]

SEQUENCE FROM N.A.  
TISSUE-PERIPHERAL BLOOD;  
MEDLINE: 95286314.  
TAKAGI S., FUJIKAWA K., IMAI T., FUKUHARA N., FUKUDOME K.,  
MINOISHI M., TSUCHIDA S., KONNO T., HIJUMA Y., YOSHIE O.,  
Identification of a highly specific surface marker of T-cell acute  
lymphoblastic leukemia and neuroblastoma as a new member of the  
transmembrane 4 superfamily.  
INT. J. CANCER 61:706-715(1995).  
-I- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.  
-I- TISSUE SPECIFICITY: NOT SOLELY EXPRESSED IN T CELLS. EXPRESSED IN  
ACUTE MYELOIDIC LEUKEMIA CELLS OF SOME PATIENTS.  
-I- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE  
ANTIGENS (TM4 SUPERFAMILY).

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way altered, modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

CC	EMBL	D10653	G285901	-
DR	EMBL	D29808	G475006	-
DR	MIM	300096	-	-
DR	PROSITE	PS00421	TM4	1
DR	PFAM	PF00335	transmembrane4	1
KV	GLYCOPROTEIN	TRANSMEMBRANE	SIGNAL-ANCHOR	
FT	DOMAIN	1	11	CYTOPLASMIC (PROBABLE)
FT	TRANSMEM	12	35	SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	DOMAIN	36	51	EXTRACELLULAR (PROBABLE)
FT	TRANSMEM	52	70	POTENTIAL
FT	DOMAIN	71	81	CYTOPLASMIC (PROBABLE)
FT	TRANSMEM	82	107	POTENTIAL
FT	DOMAIN	108	208	EXTRACELLULAR (PROBABLE)
FT	TRANSMEM	209	229	POTENTIAL
FT	DOMAIN	230	244	CYTOPLASMIC (PROBABLE)
FT	CARBOHYD	49	49	POTENTIAL
FT	CARBOHYD	150	150	POTENTIAL
FT	CARBOHYD	153	153	POTENTIAL
FT	CARBOHYD	172	172	POTENTIAL
FT	CARBOHYD	183	183	POTENTIAL
SO	SEQUENCE	244 AA	26972 MW	226CCE34B CRC32

alignment_scores:		
Quality:	280.50	Length: 245
Ratio:	1.775	Gaps: 6
Percent Similarity:	64.490	Percent Identity: 26.939

alignment\_block:

US-09-030-606-111 x A15\_HUMAN

Align seg 1/1 to: A15\_HUMAN from: 1 to: 244

[illegible]

```

seq_name: SwissProt_37:C002_HUMAN
seq_documentation_block:
ID C002_HUMAN STANDARD; PRT; 237 AA.
AC P19075;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE TUMOR-ASSOCIATED ANTIGEN CO-029.
OS TMASF3.
GN HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDE; HOMO.
RA LINNENBACH A.J.;
RA SZALA S., KASAI Y., STEPLENSKI Z., RODECK U., KORONSKI H.,
RC MOLECULAR CLONING OF cDNA FOR THE HUMAN TUMOR-ASSOCIATED ANTIGEN
RC CO-029 AND IDENTIFICATION OF RELATED TRANSMEMBRANE ANTIGENS";
RL PROC. NATL. ACAD. SCI. U.S.A. 87:6833-6837(1990).
CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: GASTRIC, COLON, RETINAL, AND PANCREATIC
CC CARCINOMAS.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE

```

```

CC ANTIGENS (TM4 SUPERFAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through
CC between the Swiss Institute of Bioinformatics and the E
CC the European Bioinformatics Institute. There are no resu
CC use by non-profit institutions as long as its content
CC modified and this statement is not removed. Usage by an
CC entities requires a license agreement (see http://www.isb-
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL; M35252; A180926; -
DR PIR; A36056; G36056.
DR MIM; 600769; -
DR PROSITE; PS00421; TM4. 1.
DR PRAM; PF00335; transmembrane4; 1.
KW GLYCOPROTEIN; ANTIGEN; TRANSMEMBRANE; SIGNAL-ANCHOR.
FT DOMAIN 1 9
FT TRANSSEM 10 33
FT DOMAIN 34 57
FT TRANSSEM 58 72
FT DOMAIN 73 83
FT TRANSSEM 84 109
FT DOMAIN 110 205
FT TRANSSEM 206 230
FT DOMAIN 231 237
FT CARBOHYD 118
SQ SEQUENCE 237 AA; 26044 MW; 6E57ECBA CRC32;

alignment_scores:
Quality: 277.50 Length: 238
Ratio: 1.790 Gaps: 7
Percent similarity: 65.126 Percent identity: 30.252

alignment_block:
US-09-030-606-111 x COO2_HUMAN

Align seg 1/1 to: COO2_HUMAN from: 1 to: 237

141 ATTAGACCATGATGATGATCCCTTCATTTGCTCATCTTTCGTGTGGTC 190
||||| ||||| ||||| ||||| ||||| ||||| |||||
8 IleuSYrSerMetPheThrPhasnPheluPheThrPleuysglYil 24
191 AGCCCTGTGGCAGTGGGACATCGTGGGTGCATCGATGATGGGACCTTTC 240
||||| ||||| ||||| ||||| ||||| ||||| |||||
24 eleuileleuAlaIleuAlaIleIleIleIleIleIleIleIleIle 40
241 TGAGATCTTGGGGCAGTCTGCTTCAGTCGATGCATGCAGTTTGTCAACGTG 290
||||| ||||| ||||| ||||| ||||| ||||| |||||
40 InAlaIlePheglYSerclunspAlaIlySerSerSeryAlaIAlaVal 56
291 GGCTACCTCCATATCGACGCGCGGGGTGGTGGCTTCCTCTGGTTTCCT 340
||||| ||||| ||||| ||||| ||||| ||||| |||||
57 Asp..IleleuIleAlaValAlaIleIleIleIleIleIleIleIle 72
341 GGGCTGCTATGGTGTCTAGACTGAGACCAAGTGTGCCCTGTGACTTCT 390
||||| ||||| ||||| ||||| ||||| ||||| |||||
72 uglYscysglYAlaIleIleYsgluserArGyCmetleuIleuPheP 89
391 TCTTCATCTCTCTCCATCTTCATCTTCAGTGTGAGTTGACGCTGTGTC 440
||||| ||||| ||||| ||||| ||||| ||||| |||||
89 heileglYleuIleuIleuIleuIleuIleuIleuIleuIleuIle 1050
441 GCCTGTGGTGCACCAACATGCGTGACACTCTCG....ACGTTGCT 484
||||| ||||| ||||| ||||| ||||| ||||| |||||
106 GlYAlaValAlaPheIYSerIYSerAspArgIleValAsnIleuIle 122
485 GGTAGTGCCTCCATCAAGAAGATATGCT...TCCAGAGAAGACTTCA 5310
||||| ||||| ||||| ||||| ||||| ||||| |||||
122 rglunsnrIhrIYleuIleuSerAlaIleIleIleIleIleIleIle 139
532 CTCAAAGTGTGAACACCAACATGAAGAAGGCTCAAGTCTGTGCTTAC 5810
||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

139 lnglualilelvalpneingluphelscyscysglyleuval 155
582 AACTATACGATTTTGGAGACTCACCCCTACTTCAAGAGACAGTGCCT 631
156 AsnGlyAla.....AlaSprtrpGlyAsnAsnPhenGlnHisTy 168
632 TCCCCCATTTGT...TGCATGACACAGCTCACCAACACCATGAAA 678
168 rProGluLeuGlyAlaCysLeuAspLys.....GlnArgp 180
679 CCTGCACCAACAAAGGCTCACGACCAAAAGTAGAGGTTGCTTCAT 728
180 roCysGlnSerTyrAsnGlyLysGlnValTyrLysGlnThrCysLieser 196
729 CAGCTTGTGTGATGACATCCGAACATGACAGTACCCGTGGTGTGTCG 778
197 PheileuAspPheleuAlaLysAsnleuilelevalilegyllese 213
779 AGCTGAATGGGGGCTCGAGCTGCGTGCATGATTGTTCATGATC 828
213 rPheGlyLeuAlaValIleGlnIleLeuGlyLeuValPheSerMetVal 230
829 TGTACTGCATCTA 842
230 eutyrcysglnile 234

seq_name: swissprot_37:A15_MOUSE
seq_documentation_block:
ID A15_MOUSE STANDARD; PRT; 244 AA.
AC 06283;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN A15 (PE31) (TALLA HOMOLOG).
GN TMAS2 OR MXS1.
OS MUS MUSCULUS (MUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-BALB/C; TISSUE-BRAIN;
RC NAGATA M., ISHAKAWA I., FUJIKAWA K., TAKAGI S., YOSHIE O.;
RL SUBMITTED (MAR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE
ANTIGENS (TM4 SUPERFAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D26483; G685221; -
DR PROSITE: PS00421; TM4; 1.
DR PRAM: PF00335; transmembrane4; 1.
KW GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL-ANCHOR.
FT DOMAIN 1 11
FT TRANSMEM 12 35 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 36 51 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
FT TRANSMEM 52 70 EXTRACELLULAR (PROBABLE).
FT DOMAIN 71 81 CYTOPLASMIC (PROBABLE).
FT TRANSMEM 82 107 POTENTIAL.
FT DOMAIN 108 208 EXTRACELLULAR (PROBABLE).
FT TRANSMEM 209 229 POTENTIAL.
FT DOMAIN 230 244 CYTOPLASMIC (PROBABLE).
FT CARBOHYD 49 49 POTENTIAL.
FT CARBOHYD 150 150 POTENTIAL.
FT CARBOHYD 153 153 POTENTIAL.
FT CARBOHYD 172 172 POTENTIAL.

```

```

FT CARBOHYD 183 183 POTENTIAL.
SQ SEQUENCE 244 AA; 26942 MW; 1014F772 CRC32;

alignment_scores:
  Quality: 273.50 Length: 245
  Ratio: 1.720 Gaps: 6
  Percent Similarity: 64.898 Percent Identity: 26.122

alignment_block:
US-09-030-606-111 x A15_MOUSE ..

Align seq 1/1 to: A15_MOUSE from: 1 to: 244

120 ACCATGACAGTTCCTACCTTCATTAAAGACCATGATGATTCCTTCATTT 169
||| .....
3 ThrLysProValIleThrCysLeuLysThrLeuLeuIleIleTyrSerPn 19
170 GCTCATCTTCTGTGTGTGTCAGACCCCTGTGGCAGTGGCATCTGGGTGT 219
.....
19 eValPheTrpIleThrGlyValIleLeuLeuAlaValGlyValTrpGlyL 36
220 CAATCGATGGGGCATCTTCTGTAAGATCTTCGGCCACACTGTCTCCACT 269
.....
36 yLeuThrLeuGlyThrTyrIleSerLeuIleAlaGlnAsnSerThr... 51
270 GCCATGCGATTTTGTCAACGTCGGGCTACTCTCATCGACGCGCGCTGT 319
||||| ||| |||||.....
52 .....AsnAlaProTyrValLeuIleGlyThrIleThr 63
320 GGTCTTGTCTGTGTGTTCTCGGGCTGCTGATGGTCTTAAGACTGAGCA 369
.....
63 rIleValIlePheGlyLeuPheGlyCysPheAlaThrCysArgGlySerp 80
370 AGTGTGCCCTCGTACCTTCTTCATCTCTCTCCATCTTCATTTGCT 419
|||
80 rOTrPheIleLysLeuTyrAlaMetPheLeuSerLeuValPheLeuAla 96
420 GAGGTTCAGCTGCTGTGTGTCGGCTTGTGTCACACCAACATGCTGAGCA 469
|||||.....
97 GluLeuValAlaGlyIleSerGlyPheValPheArgHisGluLeuLysAs 113
470 CTTCCTGACGTGTGCTGTAAGTGCCTGCATCAAGAAAGATTATGTTCC 519
...
113 pThrPheLeuAlaArgThrTyrThrAspAlaMet...GlnAsnTyrAsnGly 129
520 AGGAAGACTTCACACAGTGTGACACACACCATGAAGAGGCTCAAGTGC 569
.....
129 snAspGlnArgSerArgAlaValAspHisValGlnArgSerLeuSerCys 145
570 TGTGCTTCACCAACTATACGATTTGAGAGACTCAACCCCTACTCAAGA 619
|||||
146 CysGlyValAlaGlnAsnTyrThrAsnTrpSerSerProTyrPheLeuAs 162
620 GAAACAGTGCCTTTCCTCCCATCTGTTCACATGACCAACGTCACCAACAG 669
...
162 pHis...GlyIleProProSerCysCys.....M 171
670 CCAATGAAC...TGCACCAACAAAGGCTCACGACCAAAAGTA... 713
|||||
171 eLAsnGlnThrAspCysAsnProLeuAspLeuHisAsnLeuThrValAla 187
714 .....GAGGTGCTTCAACAGCTTTGTATGACAT 745
188 AlaThrLysValAsnGlnLysGlyCysTyrAspLeuValThrSerPhe 204
746 CCGAAGTAATGACATCACCGTGGGTGTGTGCGACGTGAATGGGGCC 795
.....
204 tGluThrAsnMetCysIleIleAlaGlyValAlaPheGlyIleAlaPheS 221
796 TCGAGCTGCTGCCATGATGTGTCCATGTATCTG 830
|||||
221 eGlnLeuIleGlyMetLeuLeuAlaCysCysLeu 232

```

seq\_name: swissprot\_37:CD37\_HUMAN

seq\_documentation\_block:

ID CD37\_HUMAN STANDARD; PRT; 281 AA.  
 AC P11049;  
 DT 01-JUL-1989 (REL. 11, CREATED)  
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE LEUKOCYTE ANTIGEN CD37.  
 GN CD37.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 89176904.  
 RA CLASSON B.J., WILLIAMS A.F., WILLIS A.C., SEED B., STAMENKOVIC I.;  
 RT "The primary structure of the human leukocyte antigen CD37, a species  
 RT homologue of the rat MRC OX-44 antigen."  
 RL J. EXP. MED. 169:1497-1502(1989).  
 RN [2]  
 RP REVISIONS.  
 RX MEDLINE: 90354767.  
 RA CLASSON B.J., WILLIAMS A.F., WILLIS A.C., SEED B., STAMENKOVIC I.;  
 RL J. EXP. MED. 172:1007-1007(1990).  
 CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: B LYMPHOCYTES.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE  
 CC ANTIGENS (TM4 SUPERFAMILY).  
 CC -1- DATABASE: NAME-PROW; NOTE-CD guide CD37 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd37.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/  
 CC or send an email to license@sdb.ch).  
 CC -----  
 CC DR EMBL: X14046; G29794; -  
 CC DR PIR: JC1500; JC1500.  
 CC DR PIR: A47629; A47629.  
 CC DR MIM: 151523; -  
 CC DR PROSITE: PS00421; TM4; 1.  
 CC DR PIR: P00035; transmembrane4; 1.  
 CC DR GLYCOPROTEIN; ANTIGEN; TRANSMEMBRANE; SIGNAL-ANCHOR.  
 CC FT DOMAIN 1 13 36 CYTOPLASMIC (PROBABLE).  
 CC FT TRANSMEM 37 59 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 CC FT TRANSMEM 60 74 EXTRACELLULAR (PROBABLE).  
 CC FT DOMAIN 60 74 POTENTIAL.  
 CC FT TRANSMEM 75 85 CYTOPLASMIC (PROBABLE).  
 CC FT TRANSMEM 86 111 POTENTIAL.  
 CC FT DOMAIN 112 241 EXTRACELLULAR (PROBABLE).  
 CC FT TRANSMEM 242 266 POTENTIAL.  
 CC FT DOMAIN 267 281 CYTOPLASMIC (PROBABLE).  
 CC FT CARBOHYD 170 170 POTENTIAL.  
 CC FT CARBOHYD 183 183 POTENTIAL.  
 CC FT CARBOHYD 188 188 POTENTIAL.  
 CC SQ SEQUENCE 281 AA; 31703 MW; 8B883F00 CRC32;

alignment\_scores:  
 Quality: 271.00 Length: 286  
 Ratio: 1.633 Gaps: 7  
 Percent Similarity: 58.042 Percent Identity: 26.573

alignment\_block:  
 US-09-030-606-111 x CD37\_HUMAN ..  
 Align seg 1/1 to: CD37\_HUMAN from: 1 to: 281

```

117 GCCACCATGACGTGCTTCACCTTCATTAAGACCATGATGATCTTCA 166
    |||  ::::::::::::::::::::|  ::::::::::::::::::::|
3  AlagInglusercysleuserleullellystyrpheuaphevalpheas 19
167 TTGTGTCATCTTTCTGTGTGTGTCAGCCCTGTGGCAGTGGCATGTGG 216
    |||  ::::::::::::::::::::|  ::::::::::::::::::::|
19  nleuphephephevalleuileuilephlephlephlephlephlephle 36
217 TGTGATCATGTGGGACATCTTTCTGAAGATCTTGGGGCCACATGTGTC 266
    ::  |||||  ::::::::::::::::::::|  ::::::::::::::::::::|
36  leulleleasplythrSerphevalSerphevalgly 48
267 AGTGCATGACGATTTGTGACGTGGGCTACTTC.....CTCATGCG 307
    ::  |||  ::::::::::::::::::::|  ::::::::::::::::::::|
49  ....leualphevalproleugInlelterSerlystalleualail 63
308 AGCCGGCGTTGTGCTTTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTA 357
    ::::::::::::::::::::|  ::::::::::::::::::::|
63  eserglyllepheThmetglyllealaleuileuglycysvalglyalal 80
358 AGACTGAGACAAAGTGGCCCTGTCGACGCTTCTTCATCTCTCTCTC 407
    ||  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
80  eulysglueuargcysleuileuglyleuayrphleglymetleuleu 96
408 ATCTTCATGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 457
    ::|  ::::::::::::::::::::|  ::::::::::::::::::::|
97  leuphealathrlnlelterleuileuglylleullelleserThrlnargal 113
458 AATGCTGACACCTTCTGACGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 507
    ::::  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
113  agInleuargSerleuargaspvalvalglyllystThrlleGlnlys 129
508 ATATGCTTCCAGAGAACTTCACT.....CAAGTGTGACACACC 548
    |||||  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
130  .lyrlelythrAsnprogluThrAlaAlaIugluserlterPaspTyr 145
549 ACCATGAAGGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 598
    ::::|  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
146  ValglInheInleuargcysglyTPrhlystyrProGlnAspTPrph 162
599 GGACTCACCTTCTCAAGAGACAGTGGC.....TTTCCC 636
    ::  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
162  eglInvalleuileuargglyAsnnglysergluAlaHlsargValProc 179
637 CATTTGTTCAATGACACAGCTCACAC..... 665
    ||  |||  ::::|  |||  ::::|  |||  ::::|  |||
179  ysSerCysTyrAsnleuSerAlaThrAsnAspSerThrlleuAsplys 195
665  .... 665
196  ValIleuProGlnleuSerAlaGleuGlyHlsleuAlaArgSerArgHl 212
666  ACAGCCATGAACCTGCACCAAGCAAGCAAGCTCAAGCAAGCAAGTAG 714
    ::::|  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
212  sSerAlaAspIleCysAlaValAlProAlaIugluserHlsleTyrArg...G 228
715  AGGCTTCAATCAACGCTTTGTATGACATCCGAACATTAAGCAGTCACCC 764
    |||||  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
228  IuglycysAlaInglyleuGlnlystPrheuhlsAsnAsnleuIleSer 244
765  GTGGGTGGTGGCAGCTGAATGGGGCTGACGCTGCTGCTGCTGCTGCTGAT 814
    ::  |||  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
245  IleValglYlleCysleuGlyValIglyleuGlnleuGlypheMetH 261
815  TGTGTCATGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTG 864
    ::::|  ::::|  |||  ::::|  |||  ::::|  |||  ::::|  |||
261  rleuSerllepheleuCysArgAsnleuAspHlsValIlyAsnArgleuA 278
865  GCCACTAC 872
    ::::|  |||
278  lAArgTyr 280

```

seq\_name: swissprot\_37:CD53\_HUMAN

```

seq_documentation_block:
ID   CD53_HUMAN          STANDARD:          PRT:          219 AA.
AC   P19397.
DT   01-NOV-1990 (REL. 16, CREATED)
DT   01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT   01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE   LEUCOCYTE SURFACE ANTIGEN CD53 (CELL SURFACE GLYCOPROTEIN CD53).
GN   CD53 OR MOX44.
OS   HOMO SAPIENS (HUMAN).
OC   EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
RN   [1]
RP   MEDLINE: 91055810.
RX   ANGELLIOVA P., VICEK C., STEFANOVA I., LIPOLDOVA M., HORESTI V.:
RT   "The human leucocyte surface antigen CD53 is a protein structurally
RL   similar to the CD37 and MRC OX-44 antigens.";
RN   [2]
RP   MEDLINE: 91079522.
RX   AMIOT M.;
RT   "Identification and analysis of cDNA clones encoding CD53. A pan-
RL   leukocyte antigen related to membrane transport proteins.";
RN   J. IMMUNOL. 145:4322-4325(1990).
CC   -1- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION IN HEMATOPOIETIC
CC       CELLS.
CC   -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.
CC   -1- TISSUE SPECIFICITY: B CELLS, MONOCYTES, MACROPHAGES, NEUTROPHILS,
CC       SINGLE (CD4 OR CD8) POSITIVE THYMOCYTES, PERIPHERAL T CELLS.
CC   -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE
CC       ANTIGENS (TM4 SUPERFAMILY).
CC   -1- DATABASE: NAME-PROW: NOTE-CD guide CD53 entry:
CC       WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd53.htm".
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@sib.ch).
CC   -----
DR   EMBL: M37033; GI80143; -.
DR   EMBL: M60871; GI80141; -.
DR   PIR: A37243; A37243.
DR   PIR: A45872; A45872.
DR   MIM: 151525; -.
DR   PROSITE: PS00421; TM4; 1.
DR   PFM: PF00335; transmembrane4; 1.
KM   GLYCOPROTEIN; ANTIGEN; TRANSMEMBRANE; SIGNAL-ANCHOR.
FT   DOMAIN 1          10
FT   TRANSMEM 11      36
FT   DOMAIN 37        54
FT   TRANSMEM 55      69
FT   DOMAIN 70        80
FT   TRANSMEM 81      106
FT   TRANSMEM 107     181
FT   DOMAIN 182      206
FT   DOMAIN 207      219
FT   CARBOHYD 129     129
FT   CARBOHYD 148     148
FT   SEQUENCE 219 AA: 24341 MW: 0FE4D13F CRC32:

alignment_scores:
Quality: 269.00      Length: 240
Ratio: 1.781        Gaps: 5
Percent Similarity: 62.917      Percent Identity: 25.833

alignment_block:
US-09-030-606-111 x CD53_HUMAN

```

```

Align seg 1/1 to: CD53_HUMAN from: 1 to: 219

123 ATGCAAGCTTCACGCTTCATTTAAGACCATGATGATCCCTTCATTTGCT 172
||||| : : : : : ||| : : : : : |||
3 MetSerSerLeuLeuLeuLeuLeuLeuLeuValLeuPhePheAsnLeuLeu 19
173 CATCTTTCTGTGGTGAGCCCTGTGGACAGTGGGACATCTGGGTGCA 222
| : : : : | : : : : | : : : : | : : : : | : : : : |
19 uphetPriLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 36
223 TCGATGGGCGATCTTTTGAAGATCTTCGGGCACTGTGTCAGTGCC 272
|| : : : ||| : : : |||
36 LeuIAsn...AsnPheGlyValLeuPhePheAsnLeuProser..... 49
273 ATGCAAGTTGTACAGCTGGGCTACTTCCATCGACGCGCGGTGTGCT 322
: : : : : ||| : : : : : |||
50 .....LeuThrLeuGlyAsnValPheValIleValGlySerIleI 63
323 CTTCGCTTGTGGTTCTGCGGCTGATGCTGAAGACTGAGACAGT 372
: : : : : ||| : : : ||| : : : |||
63 eMetValValAlaPheLeuGlyCysMetCysSerLeuLeuGlyAsnLysC 80
373 GAGCCCTCGACGCTTCTTCATCTCTCTCCATCTTCATCTTCAGT 422
|| : : : : : ||| : : : : : ||| : : : : : |||
80 yLeuLeuMetSerPhePheLeuLeuLeuLeuLeuLeuLeuLeuLeu 96
423 GTTCGACGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 472
|| : : : ||| : : : ||| : : : |||
97 ValThrLeuAlaIleLeuLeuLeuValIleGlyGlyLeuLeuAsnGly 113
473 CCGTAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 522
: : : : : ||| : : : ||| : : : |||
113 yValAlaLeuGlyLeuThrAspSerIleHisArg...TyrHisSerAsp 129
523 AAGACTTCACTCAAGTGTGACACACACACACACACACACACACAC 572
: : : : : ||| : : : ||| : : : |||
129 snSerThrLysAlaIlePhePheSerIleGlnSerPheLeuGlyCys 145
573 GCGTTCACCACTATGAGATTTTGAGGACTCACCTTACTTAAGAGAA 622
|| : : : : : ||| : : : |||
146 GlyIleAsnGlyThrSerAspTrp..... 153
623 CAGTGCCTTCCCACTCTGTGCAATGACACACACACACACACAC 672
: : : : : ||| : : : |||
154 ThrSerGlyProPheAlaSerCys..... 161
673 ATGAACCTGCACCAAGCAAAAGGCTCACGACCAAAAGTAGAGGTTGC 722
||| : : : ||| : : : |||
162 .....ProSerAspArgLysValGlyGlyCys 170
723 TTCATCAGCTTTTGTATGACATCCGACATATGACATGACCGTGGTGG 772
: : : : : ||| : : : ||| : : : |||
171 TyrAlaLysAlaArgLeuThrPheHisSerAsnPheLeuTyrIleGly 187
773 TGTGACAGCTGGAATGGGGCTGAGCTGAGCTGACATGATGTGTGCA 822
: : : : : ||| : : : ||| : : : |||
187 eIleThrIleCysValCysValIleGlyValLeuGlyMetSerPheAla 204
823 TGTATCTGACTGCAATCTA 842
: : : ||| : : : |||
204 eThrLeuAsnGlyCysGlnIle 210

seq_name: SwissProt_37:C151_MOUSE
seq_documentation_block:
ID   C151_MOUSE          STANDARD:          PRT:          253 AA.
AC   O35566.
DT   15-JUL-1998 (REL. 36, CREATED)
DT   15-JUL-1998 (REL. 36, LAST SEQUENCE UPDATE)
DT   15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE   PLATELET-ENDOTHELIAL TETRASPAN ANTIGEN 3 (PETA-3) (MEMBRANE
GLYCOPROTEIN SFA-1) (CD151 ANTIGEN).

```









```

DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PLATELET-ENDOTHELIAL TERASPAN ANTIGEN 3 (PETA-3) (GP27) (MEMBRANE
DE GLYCOPROTEIN SFA-1) (CD151 ANTIGEN).
GN CD151.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLATELET;
RX MEDLINE: 95359431.
RA FITTER S., TETAZ T.J., BERNDT M.C., ASHMAN L.K.;
RT "Molecular cloning of cDNA encoding a novel platelet-endothelial cell
RT tetra-span antigen, PETA-3."
RL BLOOD 86:1348-1355(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 96186759.
RA HASEGAWA H., UTSUNOMIYA Y., KISHIMOTO K., YANAGISAWA K., FUJITA S.;
RL "SFA-1, a novel cellular gene induced by human T-cell leukemia virus
RT type 1, is a member of the transmembrane 4 superfamily."
RL J. VIROL. 70:3258-3263(1996).
CC -1- INDUCTION: BY HTLV-1.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE
CC ANTIGENS (TMA SUPERFAMILY).
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD151 entry:
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd151.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U14650; G541613; -.
DR EMBL: D29963; D1006793; -.
DR MIM: 602243; -.
DR PROSITE: PS00421; TM4; 1.
DR PFAM: PF00335; transmembrane4.1.
KM GLYCOPROTEIN; TRANSMEMBRANE;
FT FT 1 18 SIGNAL-ANCHOR.
FT FT 19 39 CYTOPLASMIC (POTENTIAL).
FT FT 18 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT FT 18 (POTENTIAL).
FT FT 19 EXTRACELLULAR (POTENTIAL).
FT FT 40 57 POTENTIAL.
FT FT 58 78 POTENTIAL.
FT FT 79 91 POTENTIAL.
FT FT 92 112 POTENTIAL.
FT FT 113 221 EXTRACELLULAR (POTENTIAL).
FT FT 222 242 POTENTIAL.
FT FT 243 253 CYTOPLASMIC (POTENTIAL).
FT FT 159 159 POTENTIAL.
FT FT 132 132 K -> R (IN REF. 2).
FT FT 137 137 P -> S (IN REF. 2).
FT FT 137 137 CONFLICT
SQ SEQUENCE 253 AA; 28295 MW; 0E29E7ED CRC32;

```

```

alignment_scores:
    Quality: 244.50      Length: 249
    Ratio: 1.557      Gaps: 8
Percent Similarity: 63.052      Percent Identity: 28.112

alignment_block:
US-09-030-606-111 x C151_HUMAN ..

Align seg 1/1 to: C151_HUMAN from: 1 to: 253

141 ATTAGACCATGATGATCCTCTTCAATTGCTCAATCTTTCTGTGTGTC 150
    ::::| | | | | | | | | | | | | | | | | | | | | |
16 LENVYSTYLRLEULSNPHERHTYRNSYSCSYSPHETIRPLEUAIAGLYLE 32
    ..

```

```

191 AACCCGTTGGACAGGAGGGGACTGTGGGTGCATGCATGCATGGGACATCCCTTC 240
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
32 uAlaValMetAlaValAlaGlyIleThrPrlIleuAlaLeuLysSerAspTyr 49
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
241 TGAAGATCTTCGGGGCAGCTGTGTCCAGTGCAGCCATGCAGTTGTGTCAAGTG 290
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
49 LeSerIeu.....LeuAlaSerGlyThr.....TyrIleuAlaThr 60
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
291 GGCTACTCTTCATCGACAGCGGGCTGTGGTCTTGTGCTCTTGGTTGCTT 340
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
61 AlATyrIleIeuValAlaAlaGlyThrValAlaMetValThrGlyValIle 77
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
341 GGGGTGCTAGTGGCTAGACGTACAGACGAAGTGTCCCTCGTGACGTTCT 390
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
77 uGlyCysCysAlaThrPrlPheLysGluIaArgLysIleuAlaGlyLeuTyr 94
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
391 TCTTCATCTCTCTCCATCTCATCTTCATCTTCATCTTCATCTTCATCTGTGTC 440
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
94 heIleIeuLeuIleIlePheLeuLeuGluIleIleAlaSerGlyIleLeu 110
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
441 GCGTTGGTGTACACACACAATGGCTGAGCACTTCTGAGCTTGTGCTGTAGT 490
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
111 AlATyrAlaTyrTyrGluGlnIleuAsnThrGluIleuLysGluAsnLeu 127
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
491 GCGTGCATCAAGAAAGATTA.....GGTCCCAAGAGACTTCAC 534
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
127 sAsPrlMetIleThrLysArgTyrHisGlnPrlProGlyHisGluAlaValThr 144
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
535 AAGTGTGGAACACACACATGAAGAAGGGCTCAAGTGGTGGCTTCACAC 584
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
144 eAlaValAlaSerGlnIleuGlnGlnIleuPheHisCysCysGlySerHisAsn 160
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
585 TATACGGATTTTGAGGACTACCCCTACTTCAAAAGAAACAGTGC..... 629
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
161 SerGlnAspPrlArgAspSerGlnTyrPrlLeuArgSerGlnAlaIleGly 177
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
630 .....TTCCCCCATTTGTGTGCATATGACACAGTCAACACACAGCA 672
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
177 ValArgValValPrlProAspSerCysCys.....LysThrVal 189
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
673 AAGAAACCTGCACACAAAGAAAGGCTACAGCA.....AAGTA 713
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
189 AlAlaLeuCys...GlyGlnAlaArgAspHisAlaSerHisIleTyrIleVal 204
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
714 GAG.....GGTGGCTCAATCAGCTTTGTATGACATCCGAATCAATGACAGT 760
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
205 GluGlyGlyCysIleThrLysLeuGluThrPrlIleGlnGlnHisLeuArg 221
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
761 CAACGTGGGTGTGTGGCAGCTGGAATTTGGGGGCTCGAGCTGGTGGCA 810
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
221 gValIleGlyAlaValAlaGlyIleGlyIleAlaCysValGlnValPheGlu 238
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
811 TATATGTTCATGATGTATGTGTATGCAATCTCAATTAAGTGCACATC 857
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
238 eIlePheThrCysCysLeuTyrAlaArgSerLeuLysLeuGlnIleHisTyr 253
    :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
seq_name: SwissProt_37:CD53_RAT

seq_documentation_block:
ID CD53_RAT STANDARD: PRT: 218 AA.
AC P24485;
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LEUKOCYTE SURFACE ANTIGEN CD53 (CELL SURFACE GLYCOPROTEIN CD53)
DE (LEUKOCYTE ANTIGEN MRC OX-44).
GN CD53 OR OX-44.
OS RATTUS NORVEGICUS (RAT).
OC EUAROTIA; METAOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ROENTHIA; SCURROGNATHI; MORIDAE; MURINAE; RATTUS.
RN [1]
SEQUENCE FROM N.A.
RP
RX MEDLINE: 91203909.

```

RA BELLA-COSA A., LAZO P.A., BEAR S.E., TSICHLIS P.N.:  
 RT "The rat leukocyte antigen MRC OX-44 is a member of a new family of  
 RT cell surface proteins which appear to be involved in growth  
 RT regulation";  
 RL MOL. CELL. BIOL. 11:2864-2872(1991).  
 RN [2]  
 RP PRELIMINARY SEQUENCE OF 1-29.  
 RA MEDLINE; 89176904.  
 RA CLASSON B.J., WILLIAMS A.F., WILLIS A.C., SEED B., STAMENKOVIC I.:  
 RT "The primary structure of the human leukocyte antigen CD37, a species  
 RT homologue of the rat MRC OX-44 antigen";  
 RL J. EXP. MED. 169:1497-1502(1989).  
 CC -1- FUNCTION: MAY BE INVOLVED IN GROWTH REGULATION IN HEMATOPOIETIC  
 CC CELLS.  
 CC -1- SUBCELLULAR LOCATION: TYPE III MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: SPLEEN AND THYMUS, B CELLS, MONOCYTES,  
 CC MACROPHAGES, NEUTROPHILS, SINGLE (CD4 OR CD8) POSITIVE THYMOCYTES,  
 CC PERIPHERAL T CELLS.  
 CC -1- SIMILARITY: BELONGS TO THE FAMILY OF SMALL TYPE-II MEMBRANE  
 CC ANTIGENS (TM4 SUPERFAMILY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 CC EMBL; M57276; G203898; -.  
 DR PIR; A39574; A39574.  
 DR PROSITE; PS00421; TM4; 1.  
 DR PFAM; PF00335; transmembrane4; 1.  
 KW GLYCOPROTEIN; ANTIGEN; TRANSMEMBRANE; SIGNAL-ANCHOR.  
 FT INT\_MET 0 0  
 FT DOMAIN 1 9 CYTOPLASMIC (PROBABLE).  
 FT TRANSSEM 10 35 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).  
 FT DOMAIN 36 53 EXTRACELLULAR (PROBABLE).  
 FT TRANSSEM 54 68 POTENTIAL.  
 FT DOMAIN 69 79 CYTOPLASMIC (PROBABLE).  
 FT TRANSSEM 80 105 POTENTIAL.  
 FT DOMAIN 106 180 EXTRACELLULAR (PROBABLE).  
 FT TRANSSEM 181 205 POTENTIAL.  
 FT DOMAIN 206 218 CYTOPLASMIC (PROBABLE).  
 FT CARBOHYD 118 118 POTENTIAL.  
 FT CARBOHYD 128 128 POTENTIAL.  
 FT CARBOHYD 147 147 POTENTIAL.  
 SQ SEQUENCE 218 AA; 24036 MW; 8461D7CA CRC32;

alignment\_scores:  
 Quality: 242.00 Length: 243  
 Ratio: 1.603 Gaps: 6  
 Percent Similarity: 62.140 Percent Identity: 24.691

alignment\_block:  
 US-09-030-606-111 x CD53\_RAT ..

Align seg 1/1 to: CD53\_RAT from: 1 to: 218

```

123 ATGCAGTCTGCTCAGTTCATTAGACCATGATGATCCTTCATTTGCT 172
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
2 MetSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 18
173 CATCTTCTGTGTGTGTCAGCCCTGTTGGCAGTGGCATCTGGGTCAA 222
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
18 uphetrValcysgylcyscysileLeuGlyPheGlyIleHisLeuLeu 35
223 TCGATGGGCGATCTTTCGAAGATCTTGCGGCGCACTGTCGTCAGTCC 272
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
35 aIcIlnsnthr.....TyrIleLeuPheArgasn... 45
273 ATGCAGTCTGCTCAGTTCATTAGACCATGATGATCCTTCATTTGCT 322

```

```

46 LeuProPheLeuThrLeuGlyAsnValLeuValIleValGlySerIleI 62
323 CTTCGCTCTGCTGCTTCTGGGCTGCTATGTCGTAAGACTGAGACAGT 372
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
62 emetValValAlaPheLeuGlyCysMetGlySerIleGlySerIleGly 79
373 GTGCCCTGTCAGTCTTCTTCATCCTCCTCCATCTGATCTGATGCTGAG 422
||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
79 yLeuLeuMetSerPhePheValLeuLeuLeuLeuLeuLeuLeuLeu 95
423 GTTCAGCTGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 472
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
96 ValThrLeuAlaIleLeuLeuPheValTyr.....GluGly 108
473 CCGTACGCTGCTGTCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
108 sIleAsnThrLeuValAlaGluGlyLeuAsnAspSerIleGlnHisTyr 125
514 GTTCCCGAGGAGACTTCACTCAAGTGTGGAACCCACCATGAAAGGCTC 563
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
125 lSerAspAsnSerThrArgMetAlaTrpAspPheIleGlnSerGlnLeu 141
564 AAGTCTGTGCTTACCACTATACGATTTTGAGGACTACCTACTT 613
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
142 GlnCysGlyValAsnGlySerSerAspTrpIleSerGly..... 155
614 CAAAGAGAACAGTGCCTTCCCGCATCTCTGTGCAATGACAAAGTCACCA 663
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
156 .....ProPheSerSerGlyProSerGlyAla.... 164
664 ACACAGCCAAATGAACCTGCACCAAGCAAAAGGCTCAGACCAAAAGTA 713
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
165 .....AspVal 166
714 GAGGTTGCTTCAATCAGCTTTGTATGATCATCCGAACTATATGACGTAC 763
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
167 GlnGlyCysTyrLeuGlyGlyGlnAlaTrpPheHisSerAsnPheLeu 183
764 CGTGGGTGTGTGGCAGCTGGAATGGGGGCTCGACGCTGGCTGCACCA 813
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
183 rIleGlyIleValThrIleCysValCysValIleGlnValLeuGlyMet 200
814 TTGTGTCATGATCTGATCTGCAATCTA 842
||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
200 erPheAlaLeuThrLeuAsnGlyGlnIle 209

```

---

OM of: US-09-030-606-111 to: SPTREMBL\_10.\* out\_format: pfs  
Date: Sep 25, 1999 10:34 AM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL=frame+ndp.model -DEV=rlp  
-O=cpn2.1/USPTO\_SPOOL/US09030606/runat\_24091999\_171617\_29869/app-query.fasta.1  
-DB=SPTREMBL\_10 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEX=7.000 -STRAT=1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCOR=escore  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:

Query: US-09-030-606-111  
Query length: 1289  
Database: SPTREMBL\_10.\*  
Database sequences: 201082  
Database length: 61543640  
Search time (sec): 260.540000

```
Sequence      Strd Orig      ZScore      EScore Len  Documentation
SP_human:060745 + 1258.00 2040.73 3.7e-106 241 | 060745 homo sapiens (human). te
SP_human:060635 + 1254.00 2034.21 8.6e-106 241 | 060635 homo sapiens (human). te
SP_human:075954 + 310.50 486.17 4.1e-20 239 | 075954 homo sapiens (human). te
SP_human:070352 + 292.00 465.10 2.0e-18 266 | 070352 rattus norvegicus (rat). t2
SP_human:043657 + 282.00 449.50 1.6e-17 245 | 043657 rattus norvegicus (rat). t2
SP_invertebrate:019983 + 274.00 435.25 8.5e-17 282 | 019983 caenorhabditis elegans
SP_invertebrate:088429 + 270.50 430.79 1.7e-16 244 | 088429 mus musculus (mouse). cd
SP_invertebrate:070401 + 270.00 429.94 1.9e-16 245 | 070401 mus musculus (mouse). te
SP_human:060637 + 266.00 423.14 4.5e-16 253 | 060637 homo sapiens (human). te
SP_human:089118 + 263.50 419.06 7.6e-16 253 | 089118 mus musculus (mouse). p1
SP_human:060636 + 243.00 386.77 5.4e-14 222 | 060636 homo sapiens (human). te
SP_invertebrate:062745 + 241.50 383.80 7.5e-14 236 | 062745 rattus norvegicus (rat). t2
SP_invertebrate:055158 + 240.50 382.20 9.2e-14 235 | 055158 rattus norvegicus (rat). t2
SP_invertebrate:096961 + 233.00 369.51 4.4e-13 248 | 096961 geodina cydonium (spoon)
SP_human:097703 + 228.50 362.60 1.1e-12 236 | 097703 cercopithecus aethiops (p)
SP_invertebrate:096962 + 228.00 361.36 1.3e-12 248 | 096962 suberites domuncula. t
SP_human:095858 + 219.00 345.23 8.4e-12 294 | 095858 homo sapiens (human). te
SP_human:060746 + 210.50 332.17 4.9e-11 268 | 060746 homo sapiens (human). te
SP_human:060628 + 204.50 322.52 1.7e-10 264 | 060628 homo sapiens (human). te
SP_invertebrate:022495 + 199.50 313.04 5.0e-10 308 | 022495 caenorhabditis elegans
SP_invertebrate:017536 + 191.00 301.25 2.9e-09 242 | 017526 caenorhabditis elegans
SP_invertebrate:002138 + 176.00 278.70 6.5e-08 194 | 002138 caenorhabditis elegans
SP_invertebrate:018177 + 171.00 268.98 1.9e-07 233 | 018177 caenorhabditis elegans
SP_invertebrate:046101 + 171.00 268.58 1.9e-07 244 | 046101 drosophila melanogaster
SP_invertebrate:044582 + 167.50 258.08 4.2e-07 427 | 044582 caenorhabditis elegans
SP_human:075841 + 153.50 239.51 7.4e-06 260 | 075841 homo sapiens (human). te
SP_human:095859 + 151.00 234.06 1.3e-05 305 | 095859 homo sapiens (human). te
SP_invertebrate:017441 + 145.00 222.88 4.5e-05 359 | 017441 caenorhabditis elegans
SP_invertebrate:076137 + 129.00 198.75 0.0013 286 | 076137 drosophila melanogaster
SP_invertebrate:019083 + 127.50 194.00 0.0018 374 | 019083 caenorhabditis elegans
SP_human:095857 + 125.50 195.94 0.0025 204 | 095857 homo sapiens (human). te
SP_human:092971 + 122.00 182.08 0.0057 528 | 092971 sus scrofa (pig). gastr
SP_human:095552 + 122.50 166.51 0.0063 3570 | 095552 homo sapiens (human). te
SP_invertebrate:022479 + 116.50 179.63 0.0168 247 | 022479 caenorhabditis elegans
SP_human:014881 + 113.50 166.82 0.0346 622 | 014881 homo sapiens (human). te
SP_invertebrate:044420 + 110.50 170.68 0.0584 224 | 044420 schistosoma japonicum
SP_human:075851 + 107.00 140.01 0.1623 4123 | 075851 homo sapiens (human). te
SP_human:068103 + 105.50 162.61 0.1661 222 | 068103 hordobacter capsulatus (
SP_human:014335 + 102.50 150.61 0.3375 505 | 014335 homo sapiens (human). te
SP_invertebrate:076602 + 99.00 140.29 0.5064 1275 | 076602 caenorhabditis elegans
SP_invertebrate:062370 + 91.00 146.61 0.6884 417 | 062370 caenorhabditis elegans
SP_mammal:046680 + 99.00 145.07 0.7008 499 | 046680 bos taurus (bovine). tra
```

```
SP_plant:004210 + 99.00 144.88 0.7023 510 | 004210 arabidopsis thaliana
SP_fungi:094317 + 99.00 144.49 0.7056 534 | 094317 schizosaccharomyces p
SP_mammal:029070 + 98.00 147.33 0.8257 317 | 029070 sus scrofa (pig). gas
```

seq\_name: sp\_human:060745

seq\_documentation\_block:  
ID 060745 PRELIMINARY; PRT: 241 AA.

```
DT 01-AUG-1998 (TEMBREL. 07, Created)
DT 01-AUG-1998 (TEMBREL. 07, Last sequence update)
DT 01-MAY-1999 (TEMBREL. 10, Last annotation update)
DE TETRASPAN NET-1.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA RUBINSTEIN E., SERRU V., BOUCHEIX C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF065388; AAC17119.1;
DR FEM; AF00335; transmembrane4; 1;
SQ SEQUENCE 241 AA: 26301 MW: 72FSB872 CRC32;
```

alignment\_scores:  
Quality: 1258.00 Length: 241  
Ratio: 5.220 Gaps: 0  
Percent similarity: 100.000 Percent identity: 100.000

alignment\_block:  
US-09-030-606-111 x 060745

Align seg 1/1 to: 060745 from: 1 to: 241

```
123 ATGCAGCTTCACGCTTATGACATGATGATCCTTCATTTGCT 172
|||||MetGInCpHseSerPhaIleIySthrMetIleLeuPhaAsnIle 17
173 CATCTTTGTGTGTGTCGACGCTGTCGACATGCGCATCTGGGTGCA 222
|||||uIlePhLeuCySgIAlaAlaLeuAlaIValGlyIleTrrpValSerI 34
223 TCGAGGGGCACTCTTGTGAAGATCTCGGCGCACTCTCGCCAGTGC 272
|||||LeAspGlyAlaSerPhaLeuIySthIlePhGlyProLeuSerSerSerIa 50
273 ATGCAGTTGTCAACGTGGCTACTTCTCATTCGACCGCGGTGTGCT 322
|||||MetGInPhaValAsnValGlyTrrPhaLeuIleAlaIValValIa 67
323 CTTTCTCTTGGTTTCCTGGCTGCTATGCTTGAAGCTGAGACCAAGT 372
|||||lPhaIleuGlyPhaLeuGlyCySgTrrGlyAlaIySthrgInSerIySC 84
373 GCGCCCTGTGACCTCTTCTTCATCCTCCCTCTCATTTCTTATCTGAG 422
|||||GlyAlaLeuValThrPhaPhaIleuLeuLeuIlePhaIleAlaIu 100
422 GTTTCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 472
|||||ValAlaAlaAlaValAlaIleAlaValIyThrIleAlaIValIlePh 117
473 CCGTACGTTGCTGTAGTCCCTGCTCATGAAGAATTATGTTCCAGG 522
|||||eIeuThrIleuValIAlaIleIleIySthIySthIySthIySthIyS 134
523 AAGACTTACCTCACTGTGAACACACCATGAAGGCTCAAGCTGCT 572
|||||GAGCTTACCTCACTGTGAACACACCATGAAGGCTCAAGCTGCT 572
134 IuAspPhaIleuGlyAlaIySthIySthIySthIySthIySthIyS 150
573 GAGCTTACCTCACTGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 622
|||||GAGCTTACCTCACTGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCT
```

151 GlypHeThrAsnTyRThAspPheGluAspSerProTyRPhelYsgluAs 167  
 623 CAGGCGCTTCCCATTTCTGTGCAATGACGACGCTCCACCAACAGCA 672  
 167 nSerAlaPheProPheCysAsnAspAsnValThrAsnThrAla 184  
 673 ATGAACCTGCACCAAGCAAAAGGCTCAGCAACAAAGTAGAGGTTGC 722  
 184 snGluThrCysThrIysGlnIysAlaHisAspGlnIysValGluGlyCys 200  
 723 TTCATATAGCTTTGTATGACATCCGAATATGACAGTACCGGGGTGG 772  
 201 PheAsnGlnLeuLeuTyRAspIleArgThrAsnAlaValThrValGly 217  
 773 TGTGCGAGCTGGAATGGGGCTCGAGCTGCGCATGATGTGTGCA 822  
 217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSer 234  
 823 TGTATCTGTACTGCAATCTACAA 845  
 234 eTyRLeuTyRcysAsnLeuGln 241

seq\_name: sp\_human:060635

seq\_documentation\_block:  
 ID 060635 PRELIMINARY; PRT; 241 AA.  
 AC 060635;  
 DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE TSPAN-1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TODD S.C., DOCTOR V.S., LEVY S.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF054838; AAC69714.1; -;  
 DR PFM; PF00335; transmembrane4; 1.  
 SQ SEQUENCE 241 AA; 26302 MW; EF70913F CRC32;

alignment\_scores:  
 Quality: 1254.00 Length: 241  
 Ratio: 5.203 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 99.585

alignment\_block:  
 US-09-030-606-111 x 060635 ..

Align seg 1/1 to: 060635 from: 1 to: 241

123 ATGAGTGGCTTACCTTCATTAAGACCATGATGATCTCTTCATTTGCT 172  
 1 MetGlnCysPheSerPheIleIysThrMetIleLeuPheAsnLeu 17  
 173 CATTTCTGTGTGGTGGAGCCGTTGGCGATCGGCAATCGGCTGCA 222  
 17 uTlePheLeuYsgIyAlaAlaLeuLeuAlaValGlyIleThrPalSer 34  
 223 TCGATGGGCACTCTTCTGAAGATCTTGGGCACTGCTCCAGTGC 272  
 34 leaSpGlyAlaSerPheLeuYsIlePheGlyProLeuSerSerAla 50  
 273 ATGCAAGTTTGTCAAGTGGCTACTTCTCATTCAGCGCGGCTGTGCT 322  
 51 MetGlnPheValAsnValGlyTyRPhelLeuIleAlaGlyValVal 67  
 323 CTTCGCTCTGTTCTCTGGGCTGCTAATGATGCTAAAGCTAGACAGT 372  
 67 lPheAlaLeuGlyPheLeuGlyCysTyRGlYAlaIalYsThrGlnSerIysC 84

373 GTGCCCTGTGAGCTTCTTTCATCCTCCTCCATCTTCATTTGCTGAG 422  
 84 yValAlaLeuValThrPhePheIleLeuLeuLeuIlePheIleAlaGln 100  
 423 GTTCAGCTCTGTGTGCGCTTGTGTACACCAATGAGCTAGACTT 472  
 101 ValAlaAlaIleValAlaValAlaValIleValTyRThrMetIleAlaHisP 117  
 473 CCGAGCTTGTGTGTGCTGCTGCATCAAGAAAGATTAATGTTCCAG 522  
 117 eleuThrLeuLeuValAlaProAlaIleYsIysAspTyRGlYSerGln 134  
 523 AAGACTTCACTCAAGTGTGGACACCAACCATGAAGGCTCAAGTCTGT 572  
 134 lAspPheThrGlnValIleTrpAsnThrThrMetIysGlyLeuYsCys 150  
 573 GGCTTCAACCAATATACGATTTTGAGAGCTCACCTTCTCAAGAA 622  
 151 GlyPheThrAsnTyRThAspPheGluAspSerProTyRPhelYsgluAs 167  
 623 CAGTGCCTTCCCATTTCTGTGCAATGACAGCAAGTACCAACAGCCA 672  
 167 nSerAlaPheProPheCysAsnAspAsnValThrAsnThrAla 184  
 673 ATGAACCTGCACCAAGCAAAAGGCTCAGCAACAAAGTAGAGGTTGC 722  
 184 snGluThrCysThrIysGlnIysAlaHisAspGlnIysValGluGlyCys 200  
 723 TTCATATAGCTTTGTATGACATCCGAATATGACAGTACCGGGGTGG 772  
 201 PheAsnGlnLeuLeuTyRAspIleArgThrAsnAlaValThrValGly 217  
 773 TGTGCGAGCTGGAATGGGGCTCGAGCTGCGCATGATGTGTGCA 822  
 217 yValAlaAlaGlyIleGlyGlyLeuGluLeuAlaAlaMetIleValSer 234  
 823 TGTATCTGTACTGCAATCTACAA 845  
 234 eTyRLeuTyRcysAsnLeuGln 241

seq\_name: sp\_human:075954

seq\_documentation\_block:  
 ID 075954 PRELIMINARY; PRT; 239 AA.  
 AC 075954;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)  
 DE TETRASPAN NET-5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA RUBINSTEIN E., SERRU V., BOUCHEIX C.;  
 RT "New tetraspans identified in the EST database."  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF089749; AAC35859.1; -;  
 DR PFM; PF00335; transmembrane4; 1.  
 DR PROSITE; PS00421; TMA; 1.  
 SQ SEQUENCE 239 AA; 26779 MW; 4CC44BFA CRC32;

alignment\_scores:  
 Quality: 310.50 Length: 251  
 Ratio: 2.043 Gaps: 9  
 Percent Similarity: 60.558 Percent Identity: 31.873

alignment\_block:  
 US-09-030-606-111 x 075954 ..

Align seg 1/1 to: 075954 from: 1 to: 239

```

129 TCGTTCAGCTTCATTAAAGACCATGATGATCCTTCATTTGCTCATCT 178
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
5 Cysleucylcysleuylstyrmethetpheleupheasnleuilepht 21
179 TCTGTGTGTGTCAGCCCTGTTGGCAGTGCGCATCTGGGTGTCATTCGATG 228
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
21 Pheucysglycysleuleuileuglyvalglylreupheuservalsersg 38
229 GGGATCCTTCTGAAGATCTTGGCCGACATGTGTCCAGCCCATGAG 278
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
38 lnglyasnphe...Alathrpheserproserpheproser..... 50
279 TTTGTCAAGTGGGCTTCTCTCATCCAGCCGCGCTGTGTGCTTTC 328
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
51 ..leuserAlaAlasnleuVallleAlalleglythrillevalMetVa 66
329 TCTGTGTTCTCTGGGCTGCTATGCTGTCAGACTGAGAGCAAGTGTGCC 378
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
66 lthrglypheleuglycysleuglyAlallelsglunsnlyscysleul 83
379 TCGTGAGCTTCTTCATCCTCTCCATCTTCATTCATGCTGAGGTGCA 428
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
83 euleuserphephelelleuValleuVallleuLeuAlaGluleuile 99
429 GCTGCTGTGTCGCTTGTGTCACACCAATGGCTGAGCACTTCCTGAC 478
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
100 leuileuileuphephevallyrmetasplys..... 110
479 GTTGTGTGTCAGCTTCATTCAGAAAGAT..... 509
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
111 .....ValasnGlunAlaVallyslAspleuylsgluleuLeul 125
510 ..TATGTGTCCAGAGAAC.....TTCACCTCAAGTGTGAGAACCCACC 551
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
125 eutyrrhsthrGlunasnValglyleuLysAsnAlaItrpAsnIlelle 141
552 ATGAAGAGGCTCAAGTGTGCTGCTTCCACCAATTCAGATTTTGAGCA 601
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
142 GlinalaGlumetargycysglyvalthrAspyrthrAspTrp.... 156
602 CTCACCTTCTCAAGAGAACAGTGCCTTCCCATTCCTGTGTCAGATG 651
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
157 .TyrrhValleuGlunsnThrVal...Prohspargycysmetg 172
652 ACAACGTCCACCAACAGCCCAATGAACCTGCACCAACAAAGGCTAC 701
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
172 luasn.....serlnglycysglYArGAsnAlaIthrThr 183
702 GACCAAAAGTAGAGGTTGCTCAATCAGCTTTGTATGATCCGAAAC 751
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
184 ProleutrpAthrGlycyslyrleuLysVallysmetTrpAspAs 200
752 TAAATGACATCACCGTGGTGTGTCGACGTGATTTGGGGCCCTCGAGC 801
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
200 pAsnlyshlsValleuGlythrValGlyMetcyslleuileuMetGlni 217
802 TGGCTGCCATGATGTGTCATGTATCTGTACTGCATCTACATTAAGTC 851
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
217 leuaglyMetAlaphesermetThrleuphe.....GlnhlsIle 230
852 CAC 854
|||||
231 Hls 231

```

seq\_name: sp\_r0dent:070352

seq\_documentation block:

```

ID 070352 PRELIMINARY; PRT: 266 AA.
AC 070352;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE METASTASIS SUPPRESSOR HOMOLOG.

```

```

GN RAIL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC SUZUKI H., DONG J.T., GAO A.C., BARRENT J.C., ISAACS J.T.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF048882; AAC05139.1;
DR PFM; PF00335; transmembrane4; 1.
DR PROSITE; PS00421; TM4; 1.
SQ SEQUENCE 266 AA; 29487 MW; 9A1FE866 CRC32;

alignment_scores:
Quality: 292.00 Length: 266
Ratio: 1.749 Gaps: 8
Percent Similarity: 62.782 Percent Identity: 28.571

alignment_block:
US-09-030-606-111 x 070352 ..
Align seg 1/1 to: 070352 from: 1 to: 266

129 TCGTTCAGCTTCATTAAAGACCATGATGATCCTTCATTTGCTCATCT 178
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
5 CysVallysnValthrLysThrPheleupheleupheasnleuLeupheph 21
179 TCTGTGTGTGTCAGCCCTGTTGGCAGTGCGCATCTGGGTGTCATTCGATG 228
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
21 eileleuGllyAlaVallleuileuglypneglyvaltrpIleuAlaSpl 38
229 GGGATCCTTCTGAAGATCTTGGCCGACATGTGTCCAGTGCATGCATGCAG 278
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
38 ysSerSerPheIleSerValleuGlnThrSerSerSer..... 51
279 TTTGTCAAGTGGGC...TACTTCCTCATCCAGCCGCGCTGTGTGCTT 325
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
52 ..leuGlnValGlyAlaThrValPheleuileuileuileuAlaGlnVal 67
326 TGCCTGTGCTTCTCGGCTGCTATGCTGTCAGTCAAGTGTG 375
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
67 tleumetClypheleuGlycyslleGlyAlaValasnGlunValargcysL 84
376 CCTGTGTGACGCTTCTTCATCCTCTCATCTTCATTCATTCATGTCAGGTT 425
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
84 euleuGlyleuThrPheValPheleuileuileuileuileuAlaGlnVal 100
426 GCAGCTGTGTGTGCGCTTGTGTACACCAATGGCTGAGCACTTCT 475
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
101 ThrValGlnValleuPheThrPheasnAlasnlysnlysglnglume 117
476 GACGTGTGCTGTAGCTGCCATCAGAA...GATATGCTCCAGG 522
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
117 tGlyAsnThrValmetAspIlelleGlnsnThrSerAlasnIasers 134
523 AAGACTTCATCAAGTGTGAGACACCAACATGAAGGCTCAAGTGTGT 572
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
134 erserArgGlnGlnAlaItrpAspyrValGlnAlaGlnVallyscyscys 150
573 GGGTTCACCAACTATACGATTTTGAGAGCTCAACCTTACTCAAGAGAA 622
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
151 GlyTrpValserProserAsnItrpThrArgAsnProvalleuLys...As 166
623 CAGTGCCTTCCCATTCGTGTGCAT.....G 651
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
166 nserThrLysThrThrTyrrProcyssercysglulysThrlysluclua 183
652 ACAAC.....GTCAACCAACAGCCCAATGAAC 680
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :
183 spanglnleuileuVallyslsglyPhecysgluSerAspAsnSerThr 199

```





RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER R., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER L., LATREILLE P.,  
 RA LIGHTNING J., LLOYD C., MCKIBBAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PENCY C., RIKKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMER E., STADEN R., SULTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans \*;  
 RL Nature 368:32-38(1994).  
 DR EMBL; 269790; CAA93653.1; -;  
 DR PFM; PF00335; transmembrane4; 2.  
 SQ SEQUENCE 282 AA; 30942 MW; 6036A5B2 CRC32;

alignment\_scores:  
 Quality: 274.00 Length: 294  
 Ratio: 1.681 Gaps: 6  
 Percent Similarity: 55.442 Percent Identity: 24.830

## alignment\_block:

us-09-030-606-111 x 019983 ..

Align seg 1/1 to: 019983 from: 1 to: 282

```

129 TGGCTGAGCTTCATTAGACCATGATGATGCTTCTGATTCATCTT 178
    |||.....:.....:.....:.....:.....:.....:
4 CysValAsnAlaLeuAlaGlyIleValThrPheLeuPheAsnRheAlaPheTr 20
179 TCTGTGTGTGACAGCCCGTGTGGCAGTGGCAGTGGTGTCAATGATG 228
    ||| |||.....:.....:.....:.....:.....:.....:
20 PleuSerGlyValAlaValAlaPheGlyLeuGlyIleTrPheLeuPheAsp 37
229 GGGCATCTTCTGTGAAGTCTTGGGGCAGTGGTGGTGGTGGTGGTGG 278
    |||.....:.....:.....:.....:.....:.....:
37 roAlaAla...SerAspRheAlaLeuAlaHisSerThrHisProGlyAla 52
279 TTGTGCAAGTGGGCTACTTCTCTATGCGACCGGCGCTTGGTGTTC 328
    ||| |||.....:.....:.....:.....:.....:.....:
53 PheArgTrpValGlyTrpPheLeuValGlyAlaGlyAlaIleIleLe 69
329 TCTGTGTGTGACAGCCCGTGTGGCAGTGGCAGTGGTGTCAATGATG 228
    ||| |||.....:.....:.....:.....:.....:.....:
69 uValGlyTrpPheGlyCysIleGlyAlaIleTrpLysMetAsnGlnCysAla 86
379 TCGTGAAGCTTCTGTCAATCTCTCTCTATCTCTATCTCTATCTCT 428
    |||.....:.....:.....:.....:.....:.....:
86 euAlaPhePheCysCysIleLeuIleAlaPhePheLeuGlnLeuAla 102
429 GCTGTGTGTGCGCTTGTGTGACACCAATGGTGGCAGTTCCTGAC 478
    ||| |||.....:.....:.....:.....:.....:.....:
103 AlaAlaValThrLeuPheHisLysGlnGlnHisIleLysHisIleValG 119
479 GTTGTGTGTGACGCTGCTGATGAGAAAGATTATGATGCTCCAGGAAG 528
    |||.....:.....:.....:.....:.....:.....:
119 uSerSerMetLysAspThrIleArgAsnArgLysSerSerGlnThrAla 136
529 TCACATCAAGTGTGAACACACCATGAAGAAGGCTCAAGTGTGCTGC 578
    |||.....:.....:.....:.....:.....:.....:
136 heLysAspAlaPheAspRheValGlnGlnLysPheGlnCysCysLeuAl 152
579 ACCAATCTATAGGATTT.....:.....:.....:.....:.....: 607
    |||.....:.....:.....:.....:.....:.....:
153 LysThrTrpThrAspTrpLeuSerAlaArgTrpAspAlaGlnProSer 169
608 CTACTCAAGAGAAC.....:.....:.....:.....:.....:.....: 624
    |||.....:.....:.....:.....:.....:.....:
169 rGlnLeuGlnValAsnGlnGlnLysPheAlaGlyArgGlnLeuHisGly 186
625 GTGCGCTT.....:.....:.....:.....:.....:.....: 641
    |||.....:.....:.....:.....:.....:.....:

```

```

186 lValAlaPheGlyGlyAsnLysGlyThrGlyTyrglyAlaProSerSer 202
642 TGTTCATATGAC.....:.....:.....:.....:.....:.....: 667
    |||.....:.....:.....:.....:.....:.....:
203 CysLysAsnGlnHisGlyLysLeuSerTrpProAsnGlnCysGlyArgSe 219
668 AGCCATGAACCTGTGACACCAAGCAAAAGCTCAGCAGCAAAAGTAGAG 717
    |||.....:.....:.....:.....:.....:.....:
219 rPheSerGlnAlaProLeuAlaThrTrpAlaGlnPheIleAsnThrArg 236
718 GTTGTCTCAATCAGCTTTTGTATGACATCCGAACTATGAGTACCGTG 767
    |||.....:.....:.....:.....:.....:.....:
236 lYcysAlaAspAlaValTyrgIuSerValSerSerSerLeuSerIle 252
768 GGTGTGTGAGCAGCTGGAATGGGGCTGAGCTGAGCTGCGATGATGT 817
    |||.....:.....:.....:.....:.....:.....:
253 ValGlyValCysValValValLeuCysIleValGlnLeuGlyIleVal 269
818 GTCCATGATCTGTACTGCAATCTACATATGATGATGCTGCTGCTGCC 867
269 uSerMetThrLeu.....:.....:.....:.....:.....:.....: 273
868 ACTACTGCTGCCACATGGGAACTGTGAAGC 899
274 .....CysCysCysLysGlyAsnSerLysLys 282

```

seq\_name: sp\_rdent:088429

```

seq_documentation_block:
ID 088429 PRELIMINARY; PRT; 244 AA.
AC 088429;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE CELL SURFACE GLYCOPROTEIN A15.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=SPLEEN;
RA HOSOKAWA Y., SENBA E., SETO M.;
RT "Molecular cloning of a cDNA encoding mouse A15, a member of
RT transmembrane 4 superfamily and its preferential expression in brain
RT neurons.";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF052492; AAC34579.1; -;
DR PFM; PF00335; transmembrane4; 1.
SQ SEQUENCE 244 AA; 26883 MW; 096976A8 CRC32;

```

alignment\_scores:  
 Quality: 270.50 Length: 245  
 Ratio: 1.712 Gaps: 6  
 Percent Similarity: 64.490 Percent Identity: 26.122

alignment\_block:  
 us-09-030-606-111 x 088429 ..

Align seg 1/1 to: 088429 from: 1 to: 244

```

120 ACCATGAGTGTCTGAGCTTCATTAGACCATGATGATGCTTCTCAATT 169
    |||.....:.....:.....:.....:.....:.....:
3 ThrLysProValIleThrCysLeuLysThrLeuLeuIleIleTrSerPh 19
170 GCTCATCTTCTGTGTGTGACAGCCGTGTGGCAGTGGCAGTCTGGGT 219
    |||.....:.....:.....:.....:.....:.....:
19 eValPheTrpIleThrGlyValIleLeuLeuAlaValGlyValTrpGly 36
220 CAATCGATGGGCACTCTTCTGTGAAGATCTTGGGCGCACCTGTCTCAG 269
    |||.....:.....:.....:.....:.....:.....:
36 yLeuThrLeuGlyThrTrpIleSerLeuIleAlaGlnAsnSerThr... 51
    |||.....:.....:.....:.....:.....:.....:

```

```

270 GCCATGCACTTTGTCAACGTGGCTACTTCTCATTCGACGACCGCGCTGT 319
    ||||| ||| ||||| |||||
52 .....AsnAlaProTyrValLeuIleGlyThrGlyThr 63
320 GGTCTTGGCTTGGTTTCTGGGCTGCTATGGCTAAAGCTAGAGCA 369
    ||||| ||||| ||||| |||||
63 rIleValAlaPheGlyLeuPheGlyCysPheAlaThrCysArgGlySer 80
370 AGTGTGCCCCGTGACGCTTCTTCTCATCTCCCTCATCTTCTCATGT 419
    ||||| ||||| ||||| |||||
80 rorIrrPheLeuLysLeuTyrAlaMetPheLeuSerLeuValPheLeuAla 96
420 GAGGTTCACACTGCTGTGTGCGCTGGTGTACACCAATGCTGAGCA 469
    ||||| ||||| ||||| |||||
97 GluLeuValAlaGlyIleSerGlyPheValPheArgIleGlyIleLys 113
470 CTTCCTGACCTTGTGCTAGTGCCTGCCATCAGAAAGATTATGCTCC 519
    ||||| ||||| ||||| |||||
113 pIrrPheLeuArgThrTyrThrAspAlaMet...GlnAsnTyrAsnGly 129
520 AGGAAGACTTCACACTCAAGTGTGACACACACATGAAGGCTCAAGTGC 569
    ||||| ||||| ||||| |||||
129 snAspGluArgSerArgAlaValAlaPheIleGlnProSerLeuSerCys 145
570 TGTGCTTACCACTAATACGATTTTGAAGACTCACCTTACTTCAAGA 619
    ||||| ||||| ||||| |||||
146 CysGlyValGlnAsnTyrThrAsnTyrSerSerProTyrPheLeuAs 162
620 GAACAGGCTTTCCTCCCATCTTCTGCAATGACACACGTCACCAACAG 669
    ||||| ||||| ||||| |||||
162 pHis...GlyIleProTyrSerCys.....M 171
670 CCATGAAGAAC...TGCACCAAGCAAAAGGCTCAGACCAAAAGTA... 713
    ||||| ||||| ||||| |||||
171 etAsnGluThrAspCysAsnProLeuAspLeuHisAsnLeuThrValAla 187
714 .....GAGGTTGCTTCAATCAGCTTTTGTATGACAT 745
    ||||| ||||| ||||| |||||
188 AlaThrLysValAsnGlnLysGlyCysTyrAspLeuValThrSerPhe 204
746 CCACAACTAATGCAATCAGCTGCGTGTGTCGACGCTGGAATGGGGCC 795
    ||||| ||||| ||||| |||||
204 tGluThrAsnMetCylIleIleAlaGlyValAlaPheGlyIleAlaPhe 221
796 TCGAGCTGCTGCCATGATGTGTGTCATGTATCTG 830
    ||||| ||||| ||||| |||||
221 etGlnLeuIleGlyMetLeuLeuAlaLacysCysLeu 232

seq_name: sp_rdent:070401
seq_documentation_block:
ID 070401 PRELIMINARY: PRT: 245 AA.
AC 070401:
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)
DE TETRAPANIN TSPAN-6.
GN TSPAN-6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NC [1]
RA SEQUENCE FROM N.A.
RP TODD S.C., DOCTOR V.S., LEVY S.;
RL Submitted (Mar-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF051454; AAC69711.1; -.
DR PFM; PF00335; transmembrane4; 1.
SQ SEQUENCE 245 AA; 27333 MW; DCAB7D8D CRC32;

```

alignment\_scores: Quality: 270.00 Length: 241  
Ratio: 1.709 Gaps: 8

Percent Similarity: 65.560 Percent Identity: 28.631

alignment\_block:

us-09-030-606-111 x 070401 ..

Align seg 1/1 to: 070401 from: 1 to: 245

```

120 ACCATGACAGTGTTCAGCTTCAATTAAGACCATGATGCTCTTCAATT 169
    ||||| ||||| ||||| |||||
10 ThrLysProValIleThrCysLeuLysSerValLeuLeuIleTyrThr 26
170 GCTCATCTTCTGTGTGTGTCAGACCCCTGTGGCAGTGGCATCTGGGT 219
    ||||| ||||| ||||| |||||
26 etIlePheThrIleThrGlyValIleLeuLeuAlaValGlyIleThrGly 43
220 CATGCAATGGGCGCATCTTCTGAAGATCTTGGGGCATGTCTGCAT 269
    ||||| ||||| ||||| |||||
43 yVal.....SerLeuGlnAsnTyrPheSerLeuLeuAsnGlnLys 56
270 GCCATGCACTTTGTCAACGTGGGCTACTTCTCATGACGACCGCGCTGT 319
    ||||| ||||| ||||| |||||
57 Ala.....ThrAsnValProPheValIleGlyThrGlyThrVal 70
320 GGTCTTGTCTTGTGTTCTCGGCTGTATGCTGAAGTGAAGTGAAGCA 369
    ||||| ||||| ||||| |||||
70 IleIleLeuLeuGlyThrPheGlyCysPheAlaThrCysArgThrSera 87
370 AGTGTGCCCCGTGACGCTTCTTCTCATCTCTCTCATCTTCTCATGT 419
    ||||| ||||| ||||| |||||
87 IaTrpMetLeuLysLeuTyrAlaMetPheLeuThrLeuIlePheLeuVal 103
420 GAGGTTCAGCTGTGCTGTGCTGCTGTGTCACACCAATGCTGAGCA 469
    ||||| ||||| ||||| |||||
104 GluLeuValAlaAlaIleValGlyPheValPheArgHisGlyIleLysAs 120
470 CTTCCTGACGCTGTGCTGAGTGCCTGCATCAAGAAAGATTATGCTCC 519
    ||||| ||||| ||||| |||||
120 nSerPheLysSerAsnTyrGlnAsnAlaLeu...LysGluTyrAsnSera 136
520 AGGAAGACTTC...ACTCAAGTGTGGAACACACCACTGAAGGCTCAAG 566
    ||||| ||||| ||||| |||||
136 hGlyAspTyrArgSerGluAlaValAspLysIleGlnSerThrLeuHis 152
567 TGTGTGCTTACCACTAATACGATTTTGAAGACTCACCTTACTTCAA 616
    ||||| ||||| ||||| |||||
153 CysCysGlyValThrAsnTyrGlyAspTyrLysGlyThrAsnTyrTyrSe 169
617 AGAAGACAGTGCCTTTCCTCCCATCTTGTGCAATGACAAAGTCCACACA 666
    ||||| ||||| ||||| |||||
169 rGlu...ThrGlyPheProLysSerCysLys..... 179
667 CAGCCATGAACCTGCACCAAGCAAAAGCTCAGACCA...AAAGTA 713
    ||||| ||||| ||||| |||||
180 .....LeuGluGlyCysTyrProGlnArgAspAlaAspLysValAsnGlu 194
714 GAGGTTGCTTCAATCAGCTTTGTATGACATCCGAACTAATGACTGCAC 763
    ||||| ||||| ||||| |||||
195 GluGlyCysPheIleLysValAlaMetThrThrIleGlnSerGluMetCyl 211
764 CGTGTGTGTGTGTCAGCTGGAATTTGGGGCTCGAGCTGGCTGCCATGA 813
    ||||| ||||| ||||| |||||
211 lValAlaGlyIleSerPheGlyValAlaLacysPheGlnLeuIleGlyIle 228
814 TTTGTTCATGTATCTGTACTGC 836
    ||||| ||||| ||||| |||||
228 heLeuAla.....TyrCys 232

seq_name: sp_human:060637
seq_documentation_block:
ID 060637 PRELIMINARY: PRT: 253 AA.
AC 060637:
DT 01-AUG-1998 (TREMBlrel. 07, Created)

```



[illegible]

alignment\_scores:

Quality:	243.00	Length:	251
Ratio:	1.723	Gaps:	9
Percent Similarity:	56.175	Percent Identity:	28.685

```
alignment_block:
US-09-030-606-111 x 060636
```

Align seg 1/1 to: 060636 from: 1 to: 222

4

```

123 ATGCAAGCTCTTCACGTTTCATTAACACATATGATGATCTCTTCAATTTGCT 172
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
8 LeuAdgCys.....IleTyrtyleuLeuLeuIlyrhesnleuLe 21
173 CATCTTTCTGTGTGATGACACCCCGTTGGACAGTGGGCATCTGGAGTGCAA 222
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
21 uhertrpLeuLaGlyserLaValIleLaIrrhneGlyLeuTrpneArp 38
223 TCGATGGGCGATCTTTCTGTGAAGTCTTCGGGCCATGTCTCC..... 266
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 heGlyGlyAla.....IleYsgIleuSerSerGlyuAsp 49
267 AGTCCGACGACGTTTGTCACGTGGC...TACTTCTCATCCGACGGG 313
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 LysSerProGlyTrpPheTyraValGlyLeuTyraIleValGlyLaG 66
314 CGATTGTGCTTTCTCTGTGGTTTCCGCGGCTCATGTGCTAAGACTG 363
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
66 yAlaLeuMetheLaValGlyrherneGlyCysCysGlyProCysTrp 83
364 AGACGACGTGGCCCTCTGCATGTTCTTTTCATCTCTCTCCATCTTC 413
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 luserGlcysValleuGlySerPhePheTrhCysLeuLeuAlIlePhe 99
414 ATTGGTAGTGTGACGACGTCGTGGTGGCCCTGGGTGTAACACCAATAGC 463
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
100 AlaLaGlyuValTrpTrhGlyAlrPheArhPheIleGlyLysGlyAlaI 116
464 T...GAGCATCTCTGACGTTGCTGTGTAAGTGCCTGCATCAAGAAAGATT 510
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 AlIleArhTrhValGlnTrhMet..... 1233
511 ATGTTCCCGAAGAAACTTCACTCAAGTGTGGAAACCCACCATGAAGGG 560
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
124 .....TyrGlnIuLaIatYrAsnAsrTyLLeuYAsp 134
561 CTCAGTCTGTGGCTTCACCAACTATACGATTTTAGAGACTCACCTTA 610
135 ArgGly..... 136
611 CTCCAAGAGACAGTGGCTTCCCGCATTC.....T 642
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
137 ...LysGlysnGlyTrhLeuIleTrhPheProLeuGlnrherneIrc 152
643 GTTGCATGACAGCTCACACACAGCCATGAACCTGCACCAACAA 692
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
152 yScysGlyLysGlySerSerGlnGlnValGlnrProTrhCysProLysGln 168
693 AAGGTCACGACCAAAAAGTAGAGAGGTGGCTTCACTCAAGCTTTTGATGA 742
169 LeuLeu.....ArgHsLysAsnCysLleAspIuIleGlyTrpI 182
743 CATCCGAATAATGACAGTCACTGAGGTGGTGTGTGGACAGTGAATTGGG 792
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
182 eIleSerValLysLeuGlnLeuIleGlyLleValGlyIleGlyIleLaG 199
793 GCGTCGAGCTGGCTGCATGATTGTGTGCATGATGTGTACTGCATCTA 842
      :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
199 lYleuTrhIlePheGlyMetIlePheSerMetValLeuCysCysAlaIle 215
843 CAA 845
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 Arg 216

```

seq\_name: sp\_rodent:Q62745

```
seq_documentation_block:
ID      Q62745      PRELIMINARY;      PRT;      236 AA.
```

AC Q62745;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-MAY-1999 (TREMBlRel. 10, last annotation update)

DE TARGET OF THE ANTIPROLIFERATIVE ANTIBODY.

OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-CEREBRAL CORTEX;  
 RC MEDLINE: 92278609.  
 RA GEISERT E.E., MURPHY T.P., IRWIN M.H., LARJANA H.;  
 RT "A novel cell adhesion molecule, g-CAM, found on cultured rat glia.";  
 RL Neurosci. Lett. 133:262-266(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-CEREBRAL CORTEX;  
 RC MEDLINE: 93368802.  
 RA IRWIN M.H., GEISERT E.E.;  
 RT "The upregulation of a glial cell surface antigen at the astrocytic  
 scar in the rat.";  
 RL Neurosci. Lett. 154:57-60(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-CEREBRAL CORTEX;  
 RC MEDLINE: 96346153.  
 RX GEISERT E.E., YANG L., IRWIN M.H.;  
 RT "Astrocyte growth, reactivity, and the target of the  
 RT antiproliferative antibody, TAPA.";  
 RL J. Neurosci. 16:5478-5487(1996).  
 DR EMBL: U19894; AAC53103.1;  
 DR PFM: PF00335; transmembrane4; 1.  
 DR PROSITE: PS00421; TM4; 1.  
 SQ SEQUENCE 236 AA; 2588 MW; AD742C53 CRC32;

alignment\_scores:  
 Quality: 241.50 Length: 246  
 Ratio: 1.666 Gaps: 8  
 Percent Similarity: 58.943 Percent Identity: 28.862

alignment\_block:  
 US-09-030-606-111 x 062745 ..

Align seg 1/1 to: 062745 from: 1 to: 236

```

129 TGCTTCAGCTTCAATTAGACCATGATGATCCTTCAATTGTCATCTT 178
    |||  ::  |||||  :::::::::::|||||:|||||:|||||:
6  CysthrlyscystileystyrlleuLeuPheValPheAsnPheValPheTr 22
179 TCTGTGTGTGCACGCCCTGTGGCAGTGGGCATGTGGGTCAATCGATG 228
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
22 PLeuAlaGlyGlyValIleLeuGlyValAlaLeuTrpLeuArgHisAsp 39
229 GGCACATCTTTCGAAGATCTC.....GGCCACATGCTCCAGT 269
    ::::  ::::  ::::  |||  ::::  ::::
39 roGlnTrpTrpTrpLeuLeuLeuGlnLeuGlyAspLysProAlaPro 55
270 GCCATGCAAGTTTGCACAGTGGGCTACTCTTCATCCGACGCGGCTGT 319
    ::  ::  ::|||:|||||:|||||:|||||:|||||:|||||:
56 SerThrPheTrpValGlyIle...TyrIleLeuIleAlaValGlyAlaVal 71
320 GGTCTTGTGCTTGGTTCTCTGGGCTGCTATGCTGAAGACTGAGACA 369
    ::|||:|||||:|||||:|||||:|||||:|||||:|||||:
71 lMeMetPheValGlyPheLeuGlyCysTyrGlyAlaIleGlnIleuSerG 88
370 AGTGTGCCCTCGTACCTTCTCTTCATCTCCCTCCATCTTATGCT 419
    ::|||  |||  |||||  |||  |||||  |||  ::::  :::
88 lncIysLeuLeuGlyTrpPhePheTrpCysLeuValIleLeuPheAlaCys 104
420 GAGCTTCACACTGCTGTGTGCTGTGCTGTGATACACCAATG..... 461
    |||||  |||||  ::::  ::::  ::::  ::::  ::::
105 GluValAlaAlaGlyIleTrpGlyPheValAsnLysAspGlnIleAla 121
462 ....GCTGACACTTCTCTGACGTTGCTGTGATGCTCCATCAAGAAG 507
    ::::  ::::  ::::  |||||  ::::  ::::
121 sAspValIysGlnPheTrpAspGlnAlaLeuGlnGlnAlaValMetAsp 138

```

508 ATTATGTTCCCGACGAAGACTTCACTCAAGTGTGGAACACCATGAAA 557  
 || :: :::: ||| ::||| :::  
 138 sp.....AspAlaAsnAsnAlaLysAlaValIleValIleTrpPheHisGln 152  
 558 GGCCTCAAGTGTGTGGCTTACACCACTATACGATTTGAGAGCTACC 607  
 |||:||||| ||| :::: :::: :::  
 153 ThrLeuAsnLysCysGlySerAsnTrpLeuTrpThrLeuThrAla... 168  
 608 CTACTTCAAGAGAACAGTGCCTTCCCATCTCTGTGAAIAGAACAG 657  
 ::||| ||| :::: :::: :::  
 169 ....ValLeuArgAsnSerLeuCysPro.....SerSerSerAsnSer 182  
 658 TCACCAACACAGCCAAATGAACCTGCACCAAGCAAAAGCTCAGACCA 707  
 |||: ::||| ||| :::  
 182 hethrGlnLeuLeuLysGlnAspCys.....HisGln 192  
 708 AAGTAGAGGTTGCTTCAATCAGCTTTGTATGACATCGAATATGTC 757  
 |||:||||| ||| ::||| ||| :::  
 193 LysIleAspGlnLeuPheSerGlyLysLeuTrpLeuIleGlyIleAla 209  
 758 AGTCACCGTGGTGTGTGTCAGCTGGAATGGGGCTTCGAGCTGCTG 807  
 |||:||||| :: :::  
 209 aIleValAlaIleValIleMet.....IlePheG 219  
 808 CCATGATGTGTCCATGATCTGATCTGATGCAATCTACAA 845  
 |||||:||||| ||| |||:||||| :::  
 219 lueIleLeuSerMetValLeuLeuCysGlyIleArg 231

seq\_name: sp\_rodent:055158

seq\_documentation\_block:  
 ID 055158 PRELIMINARY; PRT: 235 AA.  
 AC 055158;  
 DT 01-JUN-1998 (TReMBLrel. 06, Created)  
 DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)  
 DT 01-MAY-1999 (TReMBLrel. 10, Last annotation update)  
 DE D6.1A PROTEIN.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BD IX;  
 RC MEDLINE: 98198483.  
 RX CLAAS C., SEITER S., CLAAS A., SAEVLEIVA L., SCHWAB M., ZOEGLER M.;  
 RT "Association between the rat homologue of CO-029, a  
 RT metastasis-associated tetraspanin molecule and consumption  
 RT coagulopathy";  
 RL J. Cell Biol. 141:267-280(1998).  
 DR EMBL: Y13275; CAA73724.1;  
 DR PFM: PF00335; transmembrane4; 1.  
 DR PROSITE: PS00421; TM4; 1.  
 SQ SEQUENCE 235 AA; 25580 MW; E098640B CRC32;

alignment\_scores:  
 Quality: 240.50 Length: 239  
 Ratio: 1.562 Gaps: 8  
 Percent Similarity: 64.435 Percent Identity: 28.033

alignment\_block:  
 US-09-030-606-111 x 055158 ..

Align seg 1/1 to: 055158 from: 1 to: 235

```

141 ATTAAGCATGATGATCCCTTCAATTGCTCATCTTCTGTGTGTCG 190
    ::|||  |||:|||||  |||:|||||  |||:|||||  |||:
8  LeuLysTyrSerMetPhePheAsnPheLeuPheTrpValCysGlyTrn 24
191 AGCCCTGTGGCAGTGGGCAATCTG.....GTGTCAATCGATGGGCAT 234
    ::|||:|||||  |||||  ::|||  |||||  ::
24 lLeuIleLeuGlyLeuAlaIleTrpLeuArgValSerLysAspGlyLysG 41

```

```

235 CCTTGTGAAGATCTTCGGCCACTGTCGTCAGTCGATGATGCTTGTG 284
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
41 lullele...ThrserylAspansglyhraspropheilelaval 56
285 AACGTGGGCTACTTCTCTCATCGACGCCGGCTGTGGCTTTGCTTGG 334
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
57 Aspile.....Leullelavalglyserilellemetvalleugl 70
335 TTTCCTGGGCTGCTATGCTGCTAGACTGAGACGAGTGGCCCTGTGA 384
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
70 yphleuglyCysGlyValavallysgluserArgCysmetleuul 87
385 CGTTCTTCTTCATCTCTCTCATCTTCATCTGAGTGGCTGCTGCT 434
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
87 eupheheileglyleuleuleulleuleuleuglnvalalalagly 103
435 GTGGTGGCTTGGTGTACACCAATGCTGAGCACTTCTGAGCTTGTCT 484
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
104 lleleuglyValathrPheylsergluserSerArglleleuansgluth 120
485 GGTAGTGGCCGCTCATAGAAATATATGTTCCAGAGAACTTCACTC 534
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
120 rleutyrgluanslalysleuleusergluthrSeransglualalysg 137
535 AAGTGTGAACACCACTATG.....AAAGGCTCAAGTGTGTGCG 575
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
137 luvalelnlysalametllealPhegluserglupheylscysgly 153
576 TTCACCAACTATACGATTTTGAGACTCACCTACTTCAAGAGAACAG 625
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
154 leuarg.....Pheglyalalalasprrp9lylsasn..... 164
626 TGCCTTTCCCATCTGTGTCATGACAGTCACCAACACAGCAAGT 675
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
165 ...Pheproasp.....AlalysasphlsalaglnCysthglysera 178
676 AACCTGCACCAAGCAAGGCTCAGACCCAAAGTAGAGGCTTCTTC 725
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
178 sp...CysgluserlyrasnnglygluansvalythrlythrlythrCysleu 193
726 AATCAGCTTTGTATGACATCCGAACTAATGACGTACCGCTGGGTGT 775
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
194 Serleullelyserlyrleuylslythrlellellevallelleglyl 210
776 GGCAGCTGGAATGGGGGCTGAGCTGCTGCTGATGATGTCATCT 825
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
210 ealapheglyleuallalellellellellellellellellelle 227
826 ATCTGACTGCAATCTA 842
227 alleutyrcysglnle 232
seq_name: sp_invertebrate:096961
seq_documentation_block:
ID 096961 PRELIMINARY: PRT: 248 AA.
AC 096961:
DT 01-MAY-1999 (TRENBLREL. 10. Created)
DT 01-MAY-1999 (TRENBLREL. 10. Last sequence update)
DT 01-MAY-1999 (TRENBLREL. 10. Last annotation update)
DE TERRASPININ-CD63 RECEPTOR.
GN CD63R.
OS Geodia cydonium (Sponge).
OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;
OC Astrophorida; Geodidae; Geodia.
RN
RP SEQUENCE FROM N.A.
RA WIMMER W., BOEHM M., BATEL R., SCHATTON W., FILIC Z., MUELLER W.E.G.;
RT "Initiation of aquaculture of sponges for their sustainable
RT production of bioactive metabolites in an open system: Example Geodia
RT cydonium";
RL Møl. Mar.-Biol. Biotechnol. 0:0-0(0).

```

```

DR EMBL: Y18099; CAAT7025.1; -.
KW Receptor.
SQ SEQUENCE 248 AA; 26717 MW; 78225083 CRC32;

alignment_scores:
    Quality: 233.00      Length: 261
    Ratio: 1.503        Gaps: 7
    Percent Similarity: 59.387      Percent Identity: 23.372

Alignment_block:
US-09-030-606-111 x 096961

Align seg 1/1 to: 096961 from: 1 to: 248

129 TCGTTCAGCTTATTAACACCATGATCCTCTTCAATTGCTCATCTT 178
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
10 CysglyHlsCysleuargPhehrleullelephetrpasnlelphelval 26
179 TCGTGTGTGCGAGCCCTGTGTCAGTGGGCACTGTGGTGCATCGATG 228
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
26 uleulleglyValalalaleuallalellellellellellellellelle 43
229 GGCATCCTTCTGTAAGATCTTCGGGCCACTGCTGCCAGTCCATGCG 278
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
43 snAspansntyrsergluvalthrlyserasprhralalala..... 57
279 TTGTGTCACGTGGCTACTTCTCTCATCGACGCCGCTGTGGTCTTTC 328
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
58 .....Glyalavalserllevalglyglyllellellellellelle 70
329 TCGTGTGCTCTGCTGCTGATGTCGTAAGACTGAGACAGTGTGCGC 378
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
70 yalalaleulleuglyllellellellellellellellellellellelle 87
379 TCGTGAAGTTCCTTCTTCATCTCTCTCATCTTCATGCTGAGGTCGA 428
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
87 lleglyllelellellellellellellellellellellellellellelle 103
429 GCTGCTGTGTGCTGCTGCTGTGTCACCAATGCTGAGCACTTCTTAC 478
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
104 Serglyllellellellellellellellellellellellellellellelle 120
479 GTTGTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 528
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
120 rAspansnalavalalalalellellellellellellellellellellelle 136
529 TCAGTCAAGTGTGGAACACCAACCATGAAGGCTCAAGTGTGCTGCTG 578
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
137 ..ArgArgvalilleaspalallellellellellellellellellellelle 152
579 ACCCACTATGCGATTTTGAGAGCTCACCTTCACTTCAAGAGAACAGTGC 628
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
153 AsnAsnlythrAspLythrAsp.....Arggluglugl 164
629 CTTTCCCATCTTGTG...TGCAATGACAAC..... 656
164 uilleprothrserCyslleCysaspsaspsnspclyspilyasnlysc 181
657 .....GTCCACCAACACCAACCAATGAATGAATGAC 683
181 ysilleprolleserglumetglyValaAsnserThrVal..... 193
684 ACCAAGCAAAAGCTCAGCAGCAAAAAGTAGAGGCTGCTTCAATCACT 733
194 .....TyrThrAspGlyCysArgAspSerph 202
734 TTGTATGACATCGCACTAATGACATGACGTCAGCTGCTGCTGCTGCTG 783
202 eValasprleuarggluTyrGlnleuvalalaglylallellellelle 219
784 GAATGGGGGCTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 833
    ||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```







GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:14 ; Search time 1809.22 Seconds  
(without alignments)  
1405.355 Million cell updates/sec

Title: US-09-030-606-111

Sequence: 1289  
1 AGCCAGCGCTGCTCTGCTCT.....GTTAAAAAAAAAAAAAAAAA 1289

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	621.6	48.2	697	49	AI660579
2	545	42.3	593	28	AI100799
3	518	40.2	557	51	AI708181
4	493.6	38.3	562	51	AI720370
5	478.2	37.1	607	36	AA659332
6	477.4	37.0	479	33	AA420789
7	464.8	36.1	505	34	AA468417
8	458	35.5	470	35	AA593864
9	456.8	35.4	488	27	AA037886
10	444.4	34.5	446	39	AA846269
11	438	34.0	450	42	AI075023
12	437.6	33.9	453	36	AA659579
13	434.2	33.7	496	34	AA534171
14	431.4	33.5	456	39	AA861869
15	428	33.2	440	48	AI568832
16	426.4	33.1	454	29	AA132967
17	425	33.0	425	35	AA552617
18	421	32.7	433	39	AA838513
19	421	32.7	442	44	AI276101
20	420.4	32.6	434	29	AI393647
21	420	32.6	432	45	AA161043
22	418	32.4	430	41	AI041400
23	417	32.4	497	28	AA100852
24	411.4	31.9	446	38	AA812351
25	409.6	31.8	437	35	AA58636
26	409.4	31.8	412	29	AA132677
27	402.4	31.2	426	29	AA146606
28	400	31.0	403	35	AA574382
29	399.4	31.0	429	30	AA224847
30	391	30.3	439	28	AA088704
31	388	30.1	400	50	AI670721
32	383	29.7	431	29	AA146775
33	382.8	29.7	655	50	AI673653
34	381.8	29.6	435	29	AA149408
35	381.8	29.6	444	30	AA295484
36	379.6	29.4	388	33	AA420727
37	378.2	29.3	398	35	AA557820
38	376	29.2	388	49	AI620534
39	375	29.1	399	42	AI091832
40	373	28.9	407	29	AA161042
41	373	28.9	373	34	AA468378
42	373	28.9	399	40	AA931185
43	371.8	28.8	382	34	AA468030
44	368.4	28.6	392	49	AI660655
45	368	28.5	428	28	AA113128

ALIGNMENTS

RESULT 1  
AI660579/c  
LOCUS  
DEFINITION  
IMAGE:2346245 3  
ACCESSION  
NID

AI660579 697 bp mRNA  
w668d03.x1 Scores\_Dieckgrafe-colon\_NHCD Homo sapiens CDNA clone  
sequence.  
AI660579  
94764162

EST 10-MAY-1999  
similar to TR:060635 O60635 TSPAN-1. [1] ; mRNA

Db	398	TGGGACACTGTAAAGGACGACCCCTGGCAAGCAGCACTATTGGGGAGGGGACAGATCTTA	339
QY	943	ACAAATGCTACTTGGGCGCAGAAATGAGACCTGCCCCCTTCTGCTCTCAGACTTGGGGCTAGATAG	1002
Db	338	ACAATGTCACCTTGGGCGCAGAAATGAGACCTGCCCCCTTCTGCTCTCAGACTTGGGGCTAGATAG	279
QY	1003	GGACCACTCTTTTA-GGAAATGCTGACTTCTTCCATTGGTGGGTGGATGGGTGGGG	1061
Db	278	GGACCACTCTTTTAAGGGATGCTGACTTCTTCCATTGGTGGGTGGATGGGTGGGG	219
QY	1062	GCATTCGAGACCTCTTAAGTAGAGCAGTTCGTGGCCATTCGCCAGCTATTATAACC	1121
Db	218	GCATTCGAGACCTCTTAAGTAGAGCAGTTCGTGGCCATTCGCCAGCTATTATAACC	159
QY	1122	TTGATATGCCCCCTTAGGCGCTAGTGGTATCCAGTGTCTACTGGGGGATGAGAAAG	1181
Db	158	TTGATATGCCCCCTTAGGCGCTAGTGGTATCCAGTGTCTACTGGGGGATGAGAAAG	99
QY	1182	CATTTTAAAGCTTGGGCGCTAAGTAGTAATCAGACAGAGCTCTGGGTGATGTGTAGAGCC	1241
Db	98	CATTTTAAAGCTTGGGCGCTAAGTAGTAATCAGACAGAGCTCTGGGTGATGTGTAGAGCC	39
QY	1242	ACTTCAAAATGCATATAACCTGTTCATATGTAAAAAA	1279
Db	38	ACTTCAAAATGCATATAACCTGTTCATATGTAAAAAA	1
RESULT	2		
LOCUS	AA100799/c		
DEFINITION	zm26d01.s1 Striatagene pancreas (#937208) Homo sapiens cDNA clone	EST	23-DEC-1997
ACCESSION	AA100799	IMAGE:326753	3', mRNA sequence.
NID	91647216		
VERSION	AA100799.1	GI:1647216	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
AUTHORS	1 (bases 1 to 593) Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chippelli, B., Chisose, S., Dietrich, N., Dubuque, T., Faveillo, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Merdis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rolfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R. and Warr, M.		
TITLE	Generation and analysis of 260,000 human expressed sequence tags		
JOURNAL	Genome Res. 6 (9), 807-828 (1996)		
MEDLINE	97044478		
COMMENT	On May 8, 1995 this sequence version replaced gi:800260.		
FEATURES			
source	Source		
	1. 593		
	/organism="Homo sapiens"		
	/db_xref="GDB:391812"		

```

/db_xref="taxon:9606"
/clone="IMAGE:526753"
/clone_lib="Stratagene pancreas (#937208)"
/lab_host="SOLR cells (Xanamaxin resistant)"
/notes="Organ: pancreas; Vector: plusscript SK+; Site:1
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor
sequence: 5' GATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCAGCTTTTTTTTTTTTTTTT 3'."

```

Query Match	42.3%	Score 545:	DB 28:	Length 593:
Best Local Similarity	97.8%	Prod. No.	1.1e-141:	
Matches 570; Conservative		Mismatches	11:	Indels 2; Gaps 2

Oy	693	AAGCTCACGACCCAAAAGTAAAGAGGTGGTTCATACACCTTTTGATATACATCCGAACT	752
Db	582	AAAGGCCNCCGACCAAAAAGTA -AAGATGGCTTAACTACACTTTNGATATACATCCGAACT	524
Oy	753	AATGCAGTCACCGTGGGTGGGTGG -CAAGTCGGAATGGGGGCGCTCGAGCTGGTCGCAT	811
Db	523	AATGCAGTCACCGTGGGTGGGTGGG -CAAGTCGGAATGGGGGCGCTCGAGCTGGTCGCAT	464
Oy	812	GATGTGTGCCATGTATCTGTACTGTCGAATCTACAAATAAGTCACCTTGGCTTGGCCACTA	871
Db	463	GATGTGTGCCATGTATCTGTACTGTCGAATCTACAAATAAGTCACCTTGGCTTGGCCACTA	404
Oy	872	CTGCTGCCACATGGGAACTGTGAAGAGGACCCCTGGCAAGCAGAGTGAATGGGGGAGG	931
Db	403	CTGCTGCCACATGGGAACTGTGAAGAGGACCCCTGGCAAGCAGAGTGAATGGGGGAGG	344
Oy	932	GACAGAGTCTAACAAATGTCACTTGGGCGAGAAATGAGACTGCCCTTTCGCTCCAGACTTG	991
Db	343	GACAGAGTCTAACAAATGTCACTTGGGCGAGAAATGAGACTGCCCTTTCGCTCCAGACTTG	284
Oy	992	GGGCTAGTAAAGGACCACTCTCTTTAGCCATGCCCTACTTTCCTTCATTTGGTGGGTGA	1051
Db	283	GGGCTAGTAAAGGACCACTCTCTTTAGCCATGCCCTACTTTCCTTCATTTGGTGGGTGA	224
Oy	1052	TGGGTGGGGGGGCATTTCCAGAGAGCCCTAAGGTAGCCAGGTTCTGTGGCCAAATCCCCCACTC	1111
Db	223	TGGGTGGGGGGGCATTTCCAGAGAGCCCTAAGGTAGCCAGGTTCTGTGGCCAAATCCCCCACTC	164
Oy	1112	TATTAAACCCCTTGATATATGCCCCCTAGGCCCTAAGTGGTGATCCAGTGCCTTACTAGGGGAT	1171
Db	163	TATTAAACCCCTTGATATATGCCCCCTAGGCCCTAAGTGGTGATCCAGTGCCTTACTAGGGGAT	104
Oy	1172	GAGAGAAAGGCATTTTATAGCCTGGGCATATAAGTGAATACGACAGAGCCTCTGGGTGATG	1231
Db	103	GAGAGAAAGGCATTTTATAGCCTGGGCATATAAGTGAATACGACAGAGCCTCTGGGTGATG	44
Oy	1232	TGTGAAGGCACTTCAAAATGCAATAAACCGCTTACAAATGTTAA	1274
Db	43	TGTGAAGGCACTTCAAAATGCAATAAACCGCTTACAAATGTTAA 1	

```

RESULT      3
LOCUS       AI708181.c
DEFINITION  ai708181          557 bp          mRNA
            aa99cf08.x1 Barstead colon HPRB7 Homo sapiens cDNA clone
            IMAGE:233911 3' similar to TR:060655 060635 TSPN-1. [1] ;, mRNA
            sequence.
ACCESSION   AI708181
NID         94997957
VERSION     AI708181.1  GI:4997957
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominiidae; Homo.

```

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

1 (bases 1 to 557)  
Hiller, L., Allen, M., Bowles, L., Dubuque, T., Gaisel, G., Jost, S.,  
Kritzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Mair, M.,  
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,  
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.  
WashU-NCI human EST Project  
Unpublished (1997)  
On Dec 20, 1995 this sequence version replaced g1:1133478.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu  
This clone is available royalty-free through LINTL: contact the  
IMAGE Consortium ([info@image.lnlnl.gov](mailto:info@image.lnlnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -400P from Gibco  
High quality sequence stop: 395.

```

FEATURES
SOURCE
Location/Qualifiers
1..557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19"
/cclone="IMAGE:2335911"
/cclone_1lb="Barstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note=Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; lse strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAGCGAGCGGCCGCTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' ATTTCACGTAGTAAT 3' and 5' ATTTCAGTG 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library constructed by Bob Barstead."
BASE COUNT
145 a 157 c 132 g 123 t
ORIGIN
```

Query Match	40.2%	Score 518:	DB 51:	Length 557:
Best Local Similarity	97.8%	Pred. No.3.7e-134:		
Matches 546:	Conservative 0:	Mismatches 10:	Indels 2:	Gaps 2
QY 717	GGTTCCTCAATACAGCTTTGTATACATCCGAACTAATGACAGTCACCGCTGGGTGTG	776		
Db 557	GGGTCTTCATACAGCTTTGTATACATCGAACTTAATGCCAGTCACCTGGGGTGGTGTG	498		
QY 777	GCAGCTGGAATTTGGGGGCTCGAGCTGGCTGCGCATGATTTGTCCATGTATCTAGTCC	836		
Db 497	GCAGCTGGAATTTGGGGGCTCGAGCTGGCTGCGCATGATGTGTCCATGTATCTAGTCC	438		
QY 837	AATCTACAAATAGTCACTTCTGCTCTGCCACTACTGCTGCCAATGGGAAGTGTGAAG	896		
Db 437	AATCTACAAATAGTCACTTCTGCTCTGCCACTACTGCTGCCAATGGGAAGTGTGAAG	378		
QY 897	AGGCACCTTGCCAAACACAGATGATTTGGGGGGGAGGAGACAGATCTCAACATCTACTGG	956		
Db 377	AGGCACCTTGCCAAACACAGATGATTTGGGGGGGAGGAGACAGATCTCAACATCTACTGG	319		
QY 957	GCCAAATGGAAGCTGCTTTCTGCTGCAGACTTGGGGCTAGATAGGGACACTCCTTTT	1016		
Db 318	GCCAAATGGAAGCTGCTTTCTGCTGCAGACTTGGGGCTAGATAGGGACACTCCTTTT	259		
QY 1017	A-GCGATCCTGACTTCTCTTCATTTGGTGGGTGAGTGGGTGGGGGCAATTCAGAGCCT	1075		
Db 258	AGGCGATCCTGACTTCTCTTCATTTGGTGGGTGAGTGGGTGGGGGCAATTCAGAGCCT	199		
QY 1076	CTAGGCTAGCCAGTTCTGTGGCCATTTCCCCAGTCTATTAACCTTGATATGCCCTT	1135		

Db	198	CTAAGGTGACCAAGTTCGTGTTGCCATTCGCCCATTCACATCTATTAACCCCTGATATGCCCCC	139
QY	1136	AGGCCCTAGTGTGATCCCACTGCTCTACTGTGGGGATAGAGAAAGCATTTTATAGCCTG	1195
Db	138	AGGCTTAGTGTGATCCCACTGCTCTACTGTGGGGATAGAGAAAGCATTTTATAGCCTG	79
QY	1196	GGCATTAGTGAATACAGACAGCCCTGGGGGATGTGTGAAGGCATCTCAAAATGCAT	1255
Db	78	GGCATTAGTGAATACAGACAGCCCTGGGGGATGTGTGAAGGCATCTCAAAATGCAT	19
QY	1256	AAACCTGTTACATGTTA	1273
Db	18	AAACCTGTTACATGTTA	1
RESULT	4		
AI720370/c			
LOCUS	AI720370	562 bp	MRNA
DEFINITION	aa75g12.x1 Barstead colon HPLRB7 Homo sapiens cDNA clone	EST	10-JUN-1998
	IMAGE:2334598.3	similar to TR:060635	060635 TSPAN-1. [1] ;
	sequence.		MRNA
ACCESSION	AI720370		
NID	95037626		
VERSION	AI720370.1	GI:5037626	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	Eutheria; Plimates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 562)		
AUTHORS	Hallier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,		
	Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Maira,M.,		
	Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,		
	Reising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.		
TITLE	Meshu-NCI human EST Project		
JOURNAL	Unpublished (1997)		
COMMENT	On Jun 5, 1998 this sequence version replaced gi:3189085.		

**CONTACT:** Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estewatson.wustl.edu  
 This clone is available royalty-free through LNL ; contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Possible reversed clone: similarity on wrong strand  
 Seq primer: -40UP from Gbco  
 High quality sequence stop: 411.  
**Location/Qualifiers**  
 1..562

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2334598"
/cloned_lbp="Birstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon; Vector: pT7SD-Pac (Pharmacia) with a modified polylinker; Site.1: EcoRI; Site.2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCTCGAGTGGAGCGCCGCCCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' ATTACTACTAGTAAT 3' and 5' ATTACTAGC 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7S3 vector. Library constructed by Bob Birstead."

```

	BASE COUNT	ORIGIN
Query Match	146 a	158 c
Best Local Similarity	38.3%	Score 493.6; DB 51; Length 562; Pred. NO. 2.4e-127;

[illegible]

RESULT	5
LOCUS	AA659332
DEFINITION	AA659332 607 bp mRNA EST 03-DEC-1997 n13d10.s1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:1207891, mRNA sequence.
ACCESSION	AA659332
NID	q2595486
VERSION	AA659332.1 GI:2595486
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 607) NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL	
COMMENT	On May 5, 1995 this sequence version replaced gi:798389.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/LINL at:

www.bio.llnl.gov/bdrp/image/image.html

Insert Length: 739 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 429.

## FEATURES

## SOURCE

1. 607  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="21"  
/clone="IMAGE:1207891"  
/clone\_id="NCI\_CGAP\_P12"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="vector: PAMPI0; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of DNase-treated, total cellular RNA obtained from 5,000-10,000 microdissected preneoplastic cells histologically-determined to be prostatic intraepithelial neoplasia 2 (PIN2) cells. Double-stranded cDNA was ligated to EcoRI adaptors, 5 cycles of PCR applied to the cDNA with an adaptor-specific primer, and the resulting PCR product subcloned into PAMPI0 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

BASE COUNT 152 a 134 c 168 g 153 t  
ORIGIN

Query Match 37.1%; Score 478.2; DB 36; Length 607;  
Best Local Similarity 96.4%; Pred. No. 4.9e-123;  
Matches 564; Conservative 0; Mismatches 13; Indels 8; Gaps 7;

## FEATURES

## SOURCE

Insert Length: 1109 Std Error: 0.00  
Seq primer: -28m13 rev1 ET from Amersham  
High quality sequence stop: 390.

## LOCATION/QUALIFIERS

1. 479

/organism="Homo sapiens"

/db\_xref="GDB:5923237"

/db\_xref="taxon:9606"

/clone="IMAGE:745907"

/clone\_id="NCI\_CGAP\_P11"

/sex="Male"

/dev\_stage="45 years old"

/lab\_host="DH10B"

/note="vector: PAMPI0; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with oligo(dT)17 on 50 ng of

DNase-treated, total cellular RNA obtained from

5,000-10,000 microdissected, histologically normal

prostate epithelial cells. Double-stranded cDNA was

ligated to EcoRI adaptors, 5 cycles of PCR applied to the

cDNA with an adaptor-specific primer, and the resulting

PCR product subcloned into PAMPI0 by the UDG-cloning

method (Life Technologies). Average insert size is 600

bp. NOTE: Not directionally cloned. This library was

constructed by David Krizman."

BASE COUNT 122 a 130 c 119 g 108 t  
ORIGIN

Query Match 37.0%; Score 477.4; DB 33; Length 479;  
Best Local Similarity 99.8%; Pred. No. 7.4e-123;  
Matches 478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 534 TTCAATGATATAA-CTGTACATGTTAAAAA 577  
RESULT 6  
AA420789/c 479 bp mRNA EST 16-OCT-1997  
LOCUS nc63b06.r1 NCI\_CGAP\_P1 Homo sapiens cDNA clone IMAGE:745907, mRNA  
DEFINITION sequence.  
ACCESSION AA420789  
NID 92094677  
VERSION AA420789.1 GI:2094677  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 479)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
COMMENT Unpublished (1997)  
On Sep 12, 1996 this sequence version replaced gi:1404881.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrp/image/image.html

705 CAAAAGTAGAGGGTTCCTCAATACGCTTTGTATGACATCCGAATATATGACATCACC 764  
Db 1 CAAAAGTAGAGGGTTCCTCAATACGCTTTGTATGACATCCGAATATATGACATCACC 60  
765 GTGGGTGCTGCGCAGCTGGAATGGGGCCTCGAGCTGGCTGCATATTTGTCCATG 824  
Db 61 GTGGGTGCTGCGCAGCTGGAATGGGGCCTCGAGCTGGCTGCATATTTGTCCATG 120  
825 TATCTGATCAATCTCAATATAGTCACTTCTGCTGCACTAGTGGCCACATG 884  
Db 121 TATCTGATCAATCTCAATATAGTCACTTCTGCTGCACTAGTGGCCACATG 180  
885 GGAACCTGTAAGAGCACCCTGGCAAGCAGAGTATGGGGAGGAGGAGATCTAAC 944  
Db 181 GGAACCTGTAAGAGCACCCTGGCAAGCAGAGTATGGGGAGGAGGAGATCTAAC 240  
945 AATGCTACTGGGCGAATGAGACCTGCTTGTCTCAGACTTGGGGCTAGATAGG 1004  
Db 241 AATGCTACTGGGCGAATGAGACCTGCTTGTCTCAGACTTGGGGCTAGATAGG 300  
1005 ACCACTCTTTTAGGAGAGCTGACTTCTTCATTTGGTGGTGGATGGGGGGGCA 1064  
Db 301 AACACTCTTTTAGGAGAGCTGACTTCTTCATTTGGTGGTGGATGGGGGGGCA 360  
1065 TTCCAGAGCCTCTAAGTAGCAGCTTGTGGCCATTCGCCCACTATTAAACCTTG 1124  
Db 361 TTCCAGAGCCTCTAAGTAGCAGCTTGTGGCCATTCGCCCACTATTAAACCTTG 420  
1125 ATATGCCCCCTAGAGCCTAGTGTGATCCAGTCTCTACTGGGGGATAGAGAAAGGAT 1184  
Db 421 ATATGCCCCCTAGAGCCTAGTGTGAT -CCAGTGTCTCTACTGGGGGATAGAGAA -GCAT 478  
1185 TTATAGGCTGGGCGATAGTGAATCAGACAGAGCTCTGGGCGATGTGTAGAGGCACT 1244  
Db 479 TTATAGGCT -GGCATTAAGT -AATCAGCAGAG -CTGTGGGGATGTGTAGAG -CAC 533  
1245 TCAAAATGATTAACCTGTACATATGTTAAAAA 1289

```

Db 419 CTCTGCTCTGCCACTACTGCTGCCACATGGAACTGTGAAGAGCACCCTGGCAAGCA 360
OY 914 GCAGTATGTTGGGGAGGGAGCAGGATCTAAACAATGTCACCTTTGGGCCAGAAATGGACCTGCC 973
Db 359 GCAGTATGTTGGGGAGGGAGCAGGATCTAAACAATGTCACCTTTGGGCCAGAAATGGACCTGCC 300
OY 974 CTCTGCTCTCAGACCTTGGGGCTAGATAGGAGCAGCTCCTTTTATAGCATGCTGACCTTTC 1033
Db 299 CTCTGCTCTCAGACCTTGGGGCTAGATAGGAGCAGCTCCTTTTATAGCATGCTGACCTTTC 240
OY 1034 CTTCATATGTTGGGTGATGGGTGGGGGCAATTCAGAGCCTCTAAGAGTCCAGCTTCTG 1093
Db 239 CTTCATATGTTGGGTGATGGGTGGGGGCAATTCAGAGCCTCTAAGAGTCCAGCTTCTG 180
OY 1094 TTGGCCCTTCCCGCAGCTAATTAACCTTGATATGCCCCCTAGGGCTAGTGGGATGCC 1153
Db 179 TTGGCCCTTCCCGCAGCTAATTAACCTTGATATGCCCCCTAGGGCTAGTGGGATGCC 120
OY 1154 AGTCTCTACTGGGGGATGAGAGAAAGCATTATATAGCCTGGGCATATAGTGAATCAGC 1213
Db 119 AGTCTCTACTGGGGGATGAGAGAAAGCATTATATAGCCTGGGCATATAGTGAATCAGC 60
OY 1214 AGAGCCTCTGGGTGATGATGTAGAGCAGCTTCAAAATGCATTAACCTGTTACATGTT 1272
Db 59 AGAGCCTCTGGGTGATGATGTAGAGCAGCTTCAAAATGCATTAACCTGTTACATGTT 1

RESULT 7
LOCUS AA468417 505 bp mRNA EST 14-AUG-1997
DEFINITION nc78609.r1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:783496, mRNA
sequence.
ACCESSION AA468417
NID 92194951
VERSION AA468417.1 GI:2194951
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 505)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1292183.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www.bio.lnlnl.gov/bdip/image/image.html

Insert Length: 703 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amerisham
High quality sequence stop: 442.
Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="783496"
/clone_lib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dt)17 on 50 ng of

```

```

BASE COUNT 109 a 123 c 144 g 129 t
ORIGIN
Query Match 36 1%; Score 464.8; DB 34; Length 505;
Best Local Similarity 98.2%; Pred. No. 2.5e-115;
Matches 491; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

OY 701 CGACCAAAAGTAGAGG-GTGGCTCAATCAGCTTTGTATGATCCAGTCCAGTAATGACG 759
Db 6 CGACCAAAAGTAGAGGATGATTCATCAGCTTTGTATGATCCAGTCCAGTAATGACG 65
OY 760 TCACCGTGGGTGGTGGCACTGGAA-TTGGGGCTTCGAGCTGGCTCCATGATGTG 818
Db 66 TCACCGTGGGTGGTGGCACTGGAAATTTGGGGCTTCGAGCTGGCTCCATGATGTG 125
OY 819 TCCATGATCTGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 878
Db 126 TCCATGATCTGATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 185
OY 879 CACATGGAACTGTGAAGAGCACCCTGGCAAGCAGTGTGGGGAGGGAGGAGCA 938
Db 186 CACATGGAACTGTGAAGAGCACCCTGGCAAGCAGTGTGGGGAGGGAGGAGCA 245
OY 939 TCTAACATGTCACTTGGGCAATGACCTGCTTCTGCTCCAGCTTGGGCTAG 998
Db 246 TCTAACATGTCACTTGGGCAATGACCTGCTTCTGCTCCAGCTTGGGCTAG 305
OY 999 ATAGGGACACCTCTTTAGCGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1058
Db 306 ATAGGGACACCTCTTTAGCGATGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 365
OY 1059 GGGGCAATTCAGAGCCTCTAAGGTAGCCAGTCTGTGGCCATTCGCCAGTCTATTA 1118
Db 366 GGGGCAATTCAGAGCCTCTAAGGTAGCCAGTCTGTGGCCATTCGCCAGTCTATTA 425
OY 1119 CCCTGATATGCCCCCTAGGCTAGTGTGATCCAGTCTCTACTGGGGATGAGAGAA 1178
Db 426 CCCTGATATGCCCCCTAGGCTAGTGTGATCCAGTCTCTACTGGGGATGAGAGAA 485
OY 1179 AGGCATTTATAGCTGGGGC 1198
Db 486 AGGCATTTATAGCTGGGGC 505

RESULT 8
LOCUS AA593864 470 bp mRNA EST 25-SEP-1997
DEFINITION nm19f08.s1 NCI_CGAP_Col12 Homo sapiens cDNA clone IMAGE:1084359 3',
mRNA sequence.
ACCESSION AA593864
NID 92408542
VERSION AA593864.1 GI:2408542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 470)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

```



```

Db 428 GTCCACTTGTGCTCTGACACTACTGTCGCACATGGAGACTGTGAAGACCCCTGGC 369
Oy 909 AAGCAGCAGTATGAGGAGGAGGAGACAGATCTAACATGTCTCACTTGGGC--AGATGA 967
Db 368 AAGCAGCAGTATGAGGAGGAGGAGACAGATCTAACATGTCTCACTTGGGC--AGATGA 309
Oy 968 CCTGCCCTTCTGCTCTGACACTTTGGGGCTAGATAGGAGCAGCTCTTTAGCGATGCTG 1027
Db 308 CCTGCCCTTCTGCTCTGACACTTTGGGGCTAGATAGGAGCAGCTCTTTAGCGATGCTG 249
Oy 1028 ACTTCTCTCCATGAGTGGGTGATGGGTGGGGGCGATTCAGAGCCTTAAGGTGCGCA 1087
Db 248 ACTTCTCTCCATGAGTGGGTGATGGGTGGGGGCGATTCAGAGCCTTAAGGTGCGCA 189
Oy 1088 GTTCTGTTGCCATTTCCCGCAGTCTATTAAACCTTGATATGCCCCCTAGGCTAGTGT 1147
Db 188 GTTCTGTTGCCATTTCCCGCAGTCTATTAAACCTTGATATGCCCCCTAGGCTAGTGT 129
Oy 1148 GATCCAGTCTCTACTGAGGAGGATGAGAGAAAGCATTTTATAGCCTGGGCATTAAGTGA 1207
Db 128 GATCCAGTCTCTACTGAGGAGGATGAGAGAAAGCATTTTATAGCCTGGGCATTAAGTGA 69
Oy 1208 ATCAGAGAGCCTCTGGGTGATGTATAGAGCAGCTTCAAAATGATTAACCTGTACA 1267
Db 68 ATCAGAGAGCCTCTGGGTGATGTATAGAGCAGCTTCAAAATGATTAACCTGTACA 9
Oy 1268 ATGTT 1272
Db 8 ATGTT 4

```

```

RESULT 10
AA846269/c 446 bp mRNA EST 31-DEC-1998
LOCUS a183h06.s1 Soares_parathyroid_tumor_MbHRA Homo sapiens cDNA clone
DEFINITION IMAGE:1387451 3', mRNA sequence.
ACCESSION AA846269
NID 92932409
VERSION AA846269.1 GI:2932409
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 446)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2152365.

```

```

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/btrp/image/image.html
Insert length: 1547 Std Error: 0.00
Seq primer: -40m3 fwd ET from Amersham
High quality sequence stop: 437.
Location/Qualifiers
1. 446
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="672C09:17:17011.2:21q"
/clone="IMAGE:1387451"
/clone_lib="Soares_parathyroid_tumor_MbHRA"
/tissue_type="parathyroid tumor"

```

```

/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: parathyroid gland; Vector: pRT73
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer
[5'-
TGTATCAATCTGAAGTGGAGCGCGCCAGCAATTTTTTTTTTTTTTTTTT
T-3']}, double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pRT73
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."

```

```

BASE COUNT 110 a 118 c 107 g 111 t
ORIGIN
Query Match 34.5%; Score 444.4; DB 39; Length 446;
Best Local Similarity 99.8%; Pred. No. 1.2e-113;
Matches 445; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Oy 836 CAATCTACATTAAGTCCACTTCTGCTGACACTGCTGACATGGAAGTGA 895
Db 446 CAATCTACATTAAGTCCACTTCTGCTGACACTGCTGACATGGAAGTGA 387
Oy 896 GAGGACCCCTGGCAGACAGCATGTTGGGGAGGGAGCAGATCTAACATGTCAC 955
Db 386 GAGGACCCCTGGCAGACAGCATGTTGGGGAGGGAGCAGATCTAACATGTCAC 327
Oy 956 GGGCAGATGAGACCTGCTTCTGCTCCACACTTGGGGGCTAGATAGGACCACTCTT 1015
Db 326 GGGCAGATGAGACCTGCTTCTGCTCCACACTTGGGGGCTAGATAGGACCACTCTT 267
Oy 1016 TAGCATGCTGACTTCTCTTCATTTGGTGGTGGTGGGAGGATTCAGAGCCT 1075
Db 266 TAGCATGCTGACTTCTCTTCATTTGGTGGTGGTGGGAGGATTCAGAGCCT 207
Oy 1076 CTAAAGTACCCAGTCTGTTGGCCATTCGCCAGTCTATTAAACCTTGATGCCCCCT 1135
Db 206 CTAAAGTACCCAGTCTGTTGGCCATTCGCCAGTCTATTAAACCTTGATGCCCCCT 147
Oy 1136 AGGCTAGTGGTGAATCCAGTCTCTACTGGGGATGAGAGAAAGCATTTTATAGCCTG 1195
Db 146 AGGCTAGTGGTGAATCCAGTCTCTACTGGGGATGAGAGAAAGCATTTTATAGCCTG 87
Oy 1196 GGCATTAAGTGAATCAGCAGACCTCTGGTGGATGTGTAGAGCAGCTTCAAAATGAT 1255
Db 86 GGCATTAAGTGAATCAGCAGACCTCTGGTGGATGTGTAGAGCAGCTTCAAAATGAT 27
Oy 1256 AAACCTGTACATGTATAAAAAA 1281
Db 26 AAACCTGTACATGTATAAAAAA 1

```

```

RESULT 11
A1075023/c 450 bp mRNA EST 10-NOV-1998
LOCUS a1075023.x1 NCI-CGAP_Col2 Homo sapiens cDNA clone IMAGE:1594333 3',
DEFINITION mRNA sequence.
ACCESSION A1075023
NID 93401667
VERSION A1075023.1 GI:3401667
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 450)
REFERENCE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS

```







Insert Length: 2340 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 431.

## FEATURES

**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_1fb="IMAGE:1408786"
/clone_1fb="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/notes="Vector: pT73-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
laboratories, inc., and primed with a Not I - oligo(dT)
primer [5].
TGTTACCAATCTGAGTGGAGCGCGCCCAATTTTTTTTTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5 and was
constructed by Bento Soares and M. Fatima Bonaldo."

```

BASE COUNT	115 a	123 c	110 g	108 t
ORIGIN				

Query Match	33.58;	Score 431.4;	DB 39;	Length 456;
Post (cont)	00 38;	Post 4	00-110	

OY	821	CATGTACTGTAGTACGCAMTCATAAATAAAGCAACTTGCGCTGCGCACTACTGGTGCCA	880
Dd	456	CATGTACTGTAGTACGCAMTCATAAATAAAGCAACTTGCGCTGCGCACTACTGGTGCCA	397
OY	881	CATGGGAACCTGTGTAGAAGGCCACCCTGGCAAAGCAGCATGTATTGGGGAGGGACAAGATC	940
Dd	396	CATGGGAACCTGTGTAGAAGGCCACCCT - GCAAGCAGCAGTAGTATTGGGGAGGGACAAGATC	338
OY	941	TAACAATGTCCTTTGGGGCCAGAAATGGAGCCGCCCTTGTGCTCCAACTTTGGGGCTAGAT	1000
Dd	337	TAACAATGTCCTTTGGGGCCAGAAATGGAGCCGCCCTTGTGCTCCAACTTTGGGGCTAGAT	278
OY	1001	AGGGACCACCTCCCTTTT - AGCGATGCGCTGACTTTCCTTCCATTGTTGGTGGATGGGTGGG	1059
Dd	277	AGGGACCACCTCCCTTTTGTAGGGATGCGCTGACTTTCCTTCCATTGTTGGTGGATGGGTGGG	218
OY	1060	GGGCATTCCAGAGCCCTTTAAAGTATGCCAGTTCCTGTTGCCCATCCCCAGTCTATTAAAC	1119
Dd	217	GGGCATTCCAGAGCCCTTTAAAGTATGCCAGTTCCTGTTGCCCATCCCCAGTCTATTAAAC	158
OY	1120	CCTTGATATGCCCCTTGAAGGCTCTAGTGTGTATCCCATGCTCTATAGGGGGATGAGAGAAA	1179
Dd	157	CCCTGATATGCCCCTTGAAGGCTCTAGTGTGTATCCCATGCTCTATAGGGGGATGAGAGAAA	98
OY	1180	GGCATTATTTATAGCCTGGGCACTAAAGTAAAAACACAGAGCCTCTGGGTGGATGTGTAAAG	1239
Dd	97	GGCATTATTTATAGCCTGGGCACTAAAGTAAAAACACAGAGCCTCTGGGTGGATGTGTAAAG	38
OY	1240	GCACTTCAAAATGCATTAACCTGTTTACATGTTTAAA	1276
Dd	37	GCACTTCAAAATGCATTAACCTGTTTACATGTTTAAA	1
RESULT_15	A1568832/c		
LOCUS	A1568832	440 bp	mRNA
DEFINITION	tbl16g06.x1 NCI-GCAP_Pr28 Homo sapiens cdna clone IMAGE:2118490	3'	13-APR-1999
ACCESSION	A1568832	mRNA sequence.	
NID	94532206		
VERSION	A1568832.1	GI:4532206	
KEYWORDS	EST.		
SOURCE	human.		

## ORGANISM

Homo sapiens  
Chordata; Craniata; Vertebrata; Mammalia;  
Eukaryota; Metazoa;

## REFERENCE

1 (bases 1 to 440)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>

## AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

## JOURNAL

On Feb 18, 1999 this sequence version replaced g1:4299392.

**COMMENT**

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)

Insert Length: 852      Std Error: 0.00  
Seq primer: -40UP from G1bco  
High quality sequence stop: 396.

## FEATURES

**Source**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2118490"
/clone_1lb="NCI-CGAP_Pr28"
/sex="male"
/dev_stage="adult"
/lab_host="DH10B"
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Plasmid DNA from the
normalized library NCI-CGAP_Pr22 was prepared, and ss
circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clonoids
985608-986759, 1101192-1101959, and 1217928-1220615).
Subtraction by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      109 a      107 g      103 t
ORIGIN

```

BASE COUNT  
ORIGIN

Query Match	Score	DB	Length
33.28;	428;	48;	440;
00.00;	00.00;	00.00;	00.00;

	Matches	439;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
QY	835	GCATCTACATATAGTGCACCTTCTGCCCTCTCCCACTACTGCTGCCACATGGGAACTGTGA	894							
Db	440	GCATCTACATATAGTGCACCTTCTGCCCTCTCCCACTACTGCTGCCACATGGGAACTGTGA	381							
QY	895	AGAGGCAACCCCTGGCAGAGCAGTATGTTGGGGAGGGGACAGATCTAACATGTCACTT	954							
Db	380	AGAGGCAACCCCTGGCAGAGCAGTATGTTGGGGAGGGGACAGATCTAACATGTCACTT	321							
QY	955	GGGCGCAGATGGACCTGCGCTTCTGCTCCAGACTTGGGGCTAGATAGGAGCACTCTT	1014							
Db	320	GGGCGCAGATGGACCTGCGCTTCTGCTCCAGACTTGGGGCTAGATAGGAGCACTCTT	261							
QY	1015	TTA-GCATGCGCTCACTTCTTCCATTGATGGGTGGATGGGTGGGGGGCACTCCAGAGC	1073							
Db	260	TTAAGCGATGCTCACTTCTTCCATTGATGGGTGGATGGGTGGGGGGCACTTCCAGAGC	201							
QY	1074	CTCTAAGGTACCAAGTCTGTTGCCATTTCCCCCAAGTATTATTAACCCTTATATAGCCCC	1133							
Db	200	CTCTAAGGTACCAAGTCTGTTGCCATTTCCCCCAAGTATTATTAACCCTTATATAGCCCC	141							
QY	1134	CTAAGCCTAAGTGGTATCCACAGTCTCTACTGTTGGGGATGAGAGAAAGGCATTTTATAGCC	1193							

Db 140 CTAGGCTTACTGGTGATCCAGTGTCTTACTGGGGATGAGAGAAAGCATTTTATAGCC 81  
QY 1194 TGGGCAATAGTGAATCAGCAGAGACCTCTGGTGATGTSTAGAGGCACTTCAAAATGC 1253  
Db 80 TGGGCAATAGTGAATCAGCAGAGACCTCTGGTGATGTSTAGAGGCACTTCAAAATGC 21  
QY 1254 ATAAACCTGTTACATGTTA 1273  
Db 20 ATAAACCTGTTACATGTTA 1

Search completed: September 28, 1999, 11:29:25  
Job time: 1846 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:23:04 ; Search time 2910.9 seconds  
(without alignments)  
1408.299 Million cell updates/sec

Title: US-09-030-606-111

Perfect score: 1289

Sequence: 1 AGCAGCGCTCCCTCTGCCT.....GTTAAAAAAAAAAAAAAAAAAAA 1289

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenBank:.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl1.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_st.\*  
14: gb\_sts.\*  
15: gb\_sy.\*  
16: gb\_un.\*  
17: gb\_vl.\*  
18: em\_fun.\*  
19: em\_htg.\*  
20: em\_hum1.\*  
21: em\_hum2.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_or.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sts.\*  
31: em\_sy.\*  
32: em\_un.\*  
33: em\_vl.\*  
34: gb\_htg1.\*  
35: gb\_htg2.\*  
36: gb\_in1.\*  
37: gb\_in2.\*  
38: em\_ba1.\*  
39: em\_ba2.\*  
40: em\_hum3.\*  
41: em\_hum4.\*  
42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1266	98.2	1278	11	AF065388	Homo sapi

2	1064.8	82.6	1076	11	AF054838	AF054838 Homo sapi
3	84.6	6.6	849	11	AF054841	AF054841 Homo sapi
4	83	6.4	1358	11	AF022813	AF022813 Homo sapi
5	81.8	6.3	1432	11	AF089749	AF089749 Homo sapi
6	73.2	5.7	835	12	AF052492	AF052492 Mus muscu
7	72.2	5.6	1496	9	H0MTAP1	M33680 Human 26-KD
8	71.6	5.5	1720	12	M0STAL1A	D26483 Mouse mRNA
9	70.6	5.5	711	11	AF116600	AF116600 Pan trogl
10	67.8	5.3	711	11	AF116599	AF116599 Chloroceb
11	62.8	4.9	1452	9	H0KCD53	M60871 Human cell
12	62.8	4.9	1480	9	H0KCD53GLX	M37033 Human CD53
13	62.2	4.8	1303	12	RNU19894	U19894 Rattus norv
14	60.4	4.7	1743	9	H0MA15	D10653 Homo sapien
15	60.4	4.7	1792	9	H0MMRNAB	L10723 Human (clon
16	60.4	4.7	1731	9	H0MTAL1A1	D29808 Human mRNA
17	60.4	4.7	1234	12	M0MD3MRN	X59047 M.musculus
18	59	4.6	1699	12	RAT0X44	M57276 Rat leukocy
19	57.6	4.5	1657	12	M0SC3R21A	D14883 Mouse mRNA
20	56	4.3	687	5	AR016440	AR016440 Sequence
21	56	4.3	1120	5	AR016441	AR016441 Sequence
22	56	4.3	687	5	E05732	E05732 cDNA encod1
23	56	4.3	687	5	I13743	I13743 Sequence 2
24	56	4.3	1120	5	I13744	I13744 Sequence 3
25	56	4.3	1120	9	AGMDRAP27	D10726 African gre
26	56	4.3	687	9	H0MP24CD9A	L34068 Human antiq
27	56	4.3	1120	10	H0MRP1	X60111 H.sapiens m
28	56	4.3	1192	10	H0MANTCD9	M38690 Human CD9 a
29	55.4	4.3	875	9	H0MCD63	M58485 Human lysos
30	55.4	4.3	850	9	H0MOCGB	M59907 Human melan
31	55.4	4.3	826	10	HSME491	X07982 Human mRNA
32	55.4	4.3	2049	10	HSR21MP	X53795 Human R2 mR
33	55.4	4.3	1624	10	S48196	S48196 C33 antigen
34	55.4	4.3	818	10	S93788	S93788 ocular mela
35	55.4	4.3	1607	11	HSU20770	U20770 Human metas
36	55	4.3	854	3	RAEMEA91CD	D21264 Rabbit mRNA
37	55	4.3	1130	12	M0SCD9ANT	L08115 Mus musculu
38	54.2	4.2	660	12	M0SCGCD53	X97227 M.musculus
39	54.2	4.2	612	12	M0MIGCD53A	Z16071 Mus musculu
40	53.6	4.2	38397	36	D0C8D8	A1022018 Drosophil
41	53.4	4.1	1740	12	AF049882	AF049882 Rattus no
42	53.4	4.1	41230	36	CELQ14A11	U97592 Caenorhabdi
43	52.8	4.1	669	11	AF054839	AF054839 Homo sapi
44	51.6	4.0	1819	12	D89290	D89290 Mus musculu
45	51.6	4.0	1641	12	M0U89772	U89772 Mus musculu

## ALIGNMENTS

RESULT 1

AF065388 1278 bp mRNA PRI 25-MAY-1998

LOCUS AF065388 Homo sapiens tetraspan NET-1 mRNA, complete cds.

DEFINITION AF065388

ACCESSION AF065388

ACCESION 93152700

NID

VERSION AF065388.1 GI:3152700

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1278)

AUTHORS Rubinstei n.E., Serru.V. and Bouchel x.C.

TITLE New tetraspans identified in the EST database

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1278)

AUTHORS Rubinstei n.E., Serru.V. and Bouchel x.C.

TITLE Direct Submissio n

JOURNAL Submitted (14-MAY-1998) INSERM U268, 14 av Paul Vaillant Couturier, Villejuif 94807, France

FEATURES

source 1..1278 /organism="Homo sapiens"

/db\_xref="taxon:9606"  
/clone="IMAGE ID 511239"  
122. .847  
/note="similar to Tspan-1; TM4SF"

/codon\_start=1  
/product="tetraspan NET-1"  
/protein\_id="AAC1719.1"  
/db\_xref="PID:93152701"  
/db\_xref="GI:3152701"

/translation="MOCPSFKITMMLIFNLIFLGALLAVGIWSTDSAPLKIRG  
PLSSAMFVNVGFPLIAGVYVPAISGLGCTGAKTEBCKALVTFPLILIFIAEVA  
AAVALVYITMAEHLILVPAIKDKDSODEPTQVWNTMKKCGCFNYDFED  
SPYKENSAPFPCNDVNTYNTANETCKROKRAHDKVBGCENOLDYIRINAVYGVG  
AAGIGLELAMIIVSMYLYCNIO"

BASE COUNT 280 a 335 c 329 g 334 t  
ORIGIN

Query Match 98.2%; Score 1266; DB 11; Length 1278;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1277; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```

OY 2 GCCAGGCGTCCCTTGCCTGCGCCACTCACTGCGACACCCCGAGAGCTGTTTGTCTTG 61
DB 1 GCCAGGCGTCCCTTGCCTGCGCCACTCACTGCGACACCCCGAGAGCTGTTTGTCTTG 60
OY 62 TGGAGCCTCAGAGTTCCTCTTTGAGACCTCAGTCCCAAGAGCCCTGAAGAGAGCCAC 121
DB 61 TGGAGCCTCAGAGTTCCTCTTTGAGACCTCAGTCCCAAGAGCCCTGAAGAGAGCCAC 120
OY 122 CATGAGTGTCTTCACTTCAATTAAGACATGATGATCTCTTCAATTTGCTCATCTTCT 181
DB 121 CATGAGTGTCTTCACTTCAATTAAGACATGATGATCTCTTCAATTTGCTCATCTTCT 180
OY 182 GTGTGTGTGAGCCCTGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 241
DB 181 GTGTGTGTGAGCCCTGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAGTGGGAG 240
OY 242 GAAGATCTTGGGAGCAGTGTGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 301
DB 241 GAAGATCTTGGGAGCAGTGTGCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 300
OY 302 CATGCGAGCGCGGCTGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 361
DB 301 CATGCGAGCGCGGCTGTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 360
OY 362 TTAGAGCAAGTGTGCGCTTGTGACGTTCTTTCATCCCTCCCTCATCTTTCATGCTGA 421
DB 361 TTAGAGCAAGTGTGCGCTTGTGACGTTCTTTCATCCCTCCCTCATCTTTCATGCTGA 420
OY 422 GGTTCAGAGTGTGTGTGCTGCTGTGTGACACACCAATGGGTGAGCACTTCTGTCAGTT 481
DB 421 GGTTCAGAGTGTGTGTGCTGCTGTGTGACACACCAATGGGTGAGCACTTCTGTCAGTT 480
OY 482 GCTGTAGTGTCTGCCATCAGAAAGATTATGTTCCAGAGAACTTCACTCAAGTGTG 541
DB 481 GCTGTAGTGTCTGCCATCAGAAAGATTATGTTCCAGAGAACTTCACTCAAGTGTG 540
OY 542 GAACACCAACATGAAGAGGCTCAAGTGTGCTGCTTCAACCACTATACGATTTTGAGA 601
DB 541 GAACACCAACATGAAGAGGCTCAAGTGTGCTGCTTCAACCACTATACGATTTTGAGA 600
OY 602 CTACACCTACTTCAAGAGAACAGTGTGCTTCCCATCTCTTGTGCAATGACAGCTCAC 661
DB 601 CTACACCTACTTCAAGAGAACAGTGTGCTTCCCATCTCTTGTGCAATGACAGCTCAC 660
OY 662 CACACACAGCAATGAAGCTGCACCAAGCAAAAGGCTCAGACCAAAAGAGAGGTTG 721
DB 661 CACACACAGCAATGAAGCTGCACCAAGCAAAAGGCTCAGACCAAAAGAGAGGTTG 720
OY 722 CTTCAATAGCTTTTGTATGACATCCGAATATGAGTACCGTGGTGTGTGTCAGC 781
DB 721 CTTCAATAGCTTTTGTATGACATCCGAATATGAGTACCGTGGTGTGTGTCAGC 780

```

```

OY 782 TGAATTTGGGGGCTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841
DB 781 TGAATTTGGGGGCTCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
OY 842 ACAATTAAGTCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901
DB 841 ACAATTAAGTCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
OY 902 CCCTGGCAGACAGCAGTATTTGGGGAGAGGAGCAGACATCTAACATGTCACCTTGGCCAG 961
DB 901 CCCTGGCAGACAGCAGTATTTGGGGAGAGGAGCAGACATCTAACATGTCACCTTGGCCAG 960
OY 962 AATGAGCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
DB 961 AATGAGCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
OY 1021 ATGCTGACTTCTTCTTCCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1080
DB 1021 ATGCTGACTTCTTCTTCCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1080
OY 1081 GTAGCCATTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 GTAGCCATTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
OY 1141 TAGTGTATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 TAGTGTATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
OY 1201 AAGTGAATTCAGCAGAGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
DB 1201 AAGTGAATTCAGCAGAGCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260
OY 1261 TGTACAAATGTTAAAAA 1278
DB 1261 TGTACAAATGTTAAAAA 1278

```

RESULT 2  
AF054838  
LOCUS AF054838 1076 bp mRNA PRI 03-NOV-1998  
DEFINITION Homo sapiens tetraspan TM4SF (TSPAN-1) mRNA, complete cds.  
ACCESSION AF054838  
NID 92997740  
VERSION AF054838.1 GI:2997740  
KEYWORDS  
SOURCE  
ORGANISM  
Human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominidi; Homo.  
REFERENCE  
1 (bases 1 to 1076)  
Todd, S.C., Doctor, V.S. and Levy, S.  
Sequences and expression of six new members of the  
tetraspanin/TM4SF family  
Biochim. Biophys. Acta 1399 (1), 101-104 (1998)  
JOURNAL  
MEDLINE  
REFERENCE  
2 (bases 1 to 1076)  
Todd, S.C., Doctor, V.S. and Levy, S.  
Direct Submission  
Submitted (19-MAR-1998) Medicine, Stanford, 300 Pasteur, Stanford,  
CA 94305-5115, USA  
FEATURES  
source  
1..1076  
location/Qualifiers  
gene  
1..1076  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1..1076  
CDS  
1..726  
/gene="TSPAN-1"  
/note="TSPAN-1"  
/note="tetraspanin; transmembrane 4 super family; Tspan-1"  
/codon\_start=1  
/product="tetraspan TM4SF"  
/protein\_id="AAC69714.1"

```

/db_xref="PID:g2997741"
/db_xref="GI:2997741"
/translacion="MOCFSFKITMILFNLPLFCGALLAVIGWSDGASFKIEG
PLSSAMOFVNVGVFLIAGVYFALGFAGVYCAKRESKALVFFILLIPIAFA
AAVNAVYTTMAEHFLTLVPAIKRDYSGOEFTYOVNNTMKGILCCGENTNDEED
SPFKENSAPPPCCNDVNTANETCTEOKADHOREGCFNOLLDIRNAVTVGV
AAIGGLELAAMVSVLYCNLQ"
BASE COUNT      226 a      278 c      283 g      289 t
ORIGIN

```

```

Query Match      82.6% Score 1064.8; DB 11: Length 1076;
Best Local Similarity 99.3%; Pred. No. 4.2e-298;
Matches 1069; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 123 ATGACAGTCCTGACCTTATAGACATGATGATCCTTTCATTTGCTCATCTTCTG 182
DB 1 ATGACAGTCCTGACCTTATAGACATGATGATCCTTTCATTTGCTCATCTTCTG 60
OY 183 TGTGTGTCAGCCCTGTGGCAGTGGCATCTGGGTCAATGATGGGGCATCTTCTG 242
DB 61 TGTGTGTCAGCCCTGTGGCAGTGGCATCTGGGTCAATGATGGGGCATCTTCTG 120
OY 243 AGATCTTGGGGCCACTGTGCTCCAGTGCATGCGATGTTGTCAAGTGGGCTACTTCTC 302
DB 121 AGATCTTGGGGCCACTGTGCTCCAGTGCATGCGATGTTGTCAAGTGGGCTACTTCTC 180
OY 303 ATGCGACCCGGCTTGTGGTCTTGTCTTGTCTTCTGCTGCTCATTTGCTCAAGACT 362
DB 181 ATGCGACCCGGCTTGTGGTCTTGTCTTGTCTTCTGCTGCTCATTTGCTCAAGACT 240
OY 363 GAGAGCAAGTGTGCTGCTGAGCTTCTTCACTCCCTCATCTTCAATGCTGAG 422
DB 241 GAGAGCAAGTGTGCTGCTGAGCTTCTTCACTCCCTCATCTTCAATGCTGAG 300
OY 423 GTTGCAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 482
DB 301 GTTGCAGCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
OY 483 CTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
DB 361 CTGCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
OY 543 AACACGACATGAAGGGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
DB 421 AACACGACATGAAGGGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
OY 603 TCACCTCTACTTCAAGAGAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 662
DB 481 TCACCTCTACTTCAAGAGAACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540
OY 663 AACACGACATGAAGGGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 722
DB 541 AACACGACATGAAGGGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
OY 723 TTCAATAGCTTTGTATGACATCGAATGATGATGATGATGATGATGATGATGATGATG 782
DB 601 TTCAATAGCTTTGTATGACATCGAATGATGATGATGATGATGATGATGATGATGATG 660
OY 783 GGAATTTGGGGCTTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 842
DB 661 GGAATTTGGGGCTTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
OY 843 CAATTAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 902
DB 721 CAATTAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
OY 903 CCGTGGCAAGCGAGATGTTGGGGGGGAGAGAGATCTAATCAATGCTACTTGGGGCAGA 962
DB 781 CCGTGGCAAGCGAGATGTTGGGGGGGAGAGAGATCTAATCAATGCTACTTGGGGCAGA 840
OY 963 ATGAGACCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1022

```

```

DB 841 ATGAGACCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
OY 1023 GCCGACTTTCCTTCCATGTTGGGGGATGGGTTGGGGGCTTCCAGACCTCTAAGT 1082
DB 901 TGTGACTTTCCTTCCATGTTGGGGGATGGGTTGGGGGCTTCCAGACCTCTAAGT 960
OY 1083 AGCCAGTTCTGTTGGCCCATTCGCCAGTCTATTAACCTTGATATGCCCTAGGCCCTA 1142
DB 961 AGCCAGTTCTGTTGGCCCATTCGCCAGTCTATTAACCTTGATATGCCCTAGGCCCTA 1020
OY 1143 GTGCTATCCCATGCTCTACTGGGGGATGAGAGAAAGCATTTTATGCTTGGCC 1198
DB 1021 GTGCTATCCCATGCTCTACTGGGGGATGAGAGAAAGCATTTTATGCTTGGCC 1076

RESULT 3
AF054841      849 bp      mRNA      PRI      03-NOV-1998
LOCUS
DEFINITION   Homo sapiens telraspan TW4SF (TSPAN-4) mRNA, complete cds.
ACCESSION   AF054841
NID         92997746
VERSION      AF054841.1 GI:2997746
KEYWORDS
SOURCE
ORGANISM     human.
              Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
              Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE    1 (bases 1 to 849)
              Todd,S.C., Doctor,V.S. and Levy,S.
              Sequences and expression of six new members of the
              telraspan/TW4SF family
              Blochim. Biophys. Acta 1399 (1), 101-104 (1998)
JOURNAL
MEDLINE
89390278
REFERENCE    2 (bases 1 to 849)
              Todd,S.C., Doctor,V.S. and Levy,S.
              Direct Submision
              Submitted (19-MAR-1998) Medicine, Stanford, 300 Pasteur, Stanford,
              CA 94305-5115, USA
FEATURES
source
location/Qualifiers
      1..849
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      1..849
      /gene="TSPAN-4"
      /note="telraspan; transmembrane 4 super family; MAG-2;
      Tspan-4"
      /product="telraspan TW4SF"
      /protein_id="AAC69717.1"
      /db_xref="PID:g2997747"
      /db_xref="GI:2997747"
      /translacion="MARACLOAVKYLMEFNLFLMLGCGVLGVIWLAATGSEFATL
      SSESFSLSANLILITIGAFVMAIGFVGAIKXENKCLITFEILLVLEAITAI
      LPAAYVDXKIDRAOQDLKRGILHYGQGVAVGLTANSIITOTDERCGGVNFDMEFVY
      NTRVVDSCLESSESCGLHAPETMKAPCYETVKKVWLDENLAVGIFGLCTALVOIL
      GLTEFATMTCQVVKADTYCA"
BASE COUNT      132 a      284 c      249 g      184 t
ORIGIN

```

```

Query Match      6.6% Score 84.6; DB 11: Length 849;
Best Local Similarity 53.7%; Pred. No. 5.8e-14;
Matches 252; Conservative 0; Mismatches 199; Indels 18; Gaps 3;

```

```

OY 129 TGTTCAGCTTCAATAGACATGATGATGCTCTCAATTGCTCATCTTCTGTGTGGT 188
DB 131 TGTTCAGGCGCGTCAAGTACCTCATCTTCTGCTTCAACCTGCTTCTGCTGGGAGGC 190
OY 189 GCAGCCCTGTGGCAGTGGGATCTGGGTCAATGATGGGCAATCTTCTGAAATC 248
DB 191 TGTGCGTGTGCTGGGTGTGGCAGTCTGCT-----GGCGCCACACAGGGGAGC 238

```





```

BASE COUNT      296 a      410 c      405 g      321 t
ORIGIN

```

LOCUS	HUMTAPAL	1496 bp	mRNA	PRI	20-NOV-1992
DEFINITION	Human 26-kDa cell surface protein TAPA-1 mRNA, complete cds.				
ACCESSION	M33680				
NID	9338677				
VERSION	M33680.1	GI:338677			
KEYWORDS	26-kDa cell surface protein TAPA-1; target of antiproliferative antibody.				
SOURCE	Human cell line OCI-LY8, CDNA to mRNA, clones 7-3 and 8-1.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	Oren R., Takehashi S., Doss C., Levy R. and Levy S.				
TITLE	TAPA-1, the target of an anti-proliferative antibody, defines a new family of transmembrane proteins				
JOURNAL	Mol. Cell. Biol. 10, 4007-4015 (1990)				
MEDLINE	90318365				
COMMENT	Drift entry and computer readable sequence for [1] kindly submitted by S. Levy, 10-APR-1990, for release after publication.				
FEATURES	Location/Qualifiers				
Source	1. 1496				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
CDS	239..949				
	/note="26-kDa cell surface protein TAPA-1"				
	/codon_start=1				
	/protein_id="AA36663.1"				
	/db_xref="PID:g338678"				
	/db_xref="GI:338678"				
	/translation="MGVEGCKIKYLLFENFVFWLGGVILGVALMLRHPDTNMLYLEDKRPAPNTFVGVGYLLIAYGAVMFPVGLGCGAGDSOCILPFTCYVILFACFPAAGICGVFNKDQIAKDYKQYTDQALQAVDDNNKAKVIFHEFLDCCGSSFTLWTTSTVLNNLCIPSGSSNIISNLFEDCKIDIDFSGKLYLIGIAIYVAVIMFEETLIMVLLCCGIRNRSYV"				
BASE COUNT	polya_signal 257 a 504 c 413 g 322 t				
ORIGIN	1455..1460				
Query Match	5.6%; Score 72.2; DB 9; Length 1496;				
Best Local Similarity	49.9%; Pred. No. 2.5e-10;				
Matches 244; Conservative	0; Mismatches 233; Indels 12; Gaps 2;				
129	TGCTTCACCTTCATTAAACCATGATGATCTCTTCAATTGCTCATCTTCTGTGTGGT	188			
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111					
254	TGCACCAAGTGCATCAATACCTCTCTCGCTTCATATTTCTTCTGTGCGTGGCTGA	313			
189	GCAGCCCTGTGGCATGTGGGCATCTGGGGGTGAATGATGGG-----GCATCTTTCTG	242			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					
314	GCGGATCTCTGGGTGTGTGACCTGTGGGCTCCGCCATGACCCGACACCAACCACTCTCG	373			
243	AAGATCTTCGGGCCACTCTGTCACAGTGCATGATTTGTTCAGTGGGCTACTTCTGC	302			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					
374	TATCTGGAGCTGGGAGACAGAGCCGCCCAACACTTCTATGTAGGCACTCAATCTCTC	433			
303	ATCGCAGCCGGCGTTGTGTCTTCTCTCTTGTGTTTCTTGCGCTGCTATGTGTCCTAAGACT	362			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					
434	ATTCGTTGGGGGGCTGTATGATGTTTCTGTGGCTCTTGCGCTGTACGGGGCCATCCAG	493			
363	GAGAGCAGTGTGGCCCTGTGTGACGTTCTTCTTCATCTCCCTCCATCTCTCACTTCTGAG	422			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					
494	GATTCACAGTGTCTGTGGGAGCGTTCTTCACCTGCTGTGATCTCTGTTGCCGTGAG	553			
423	GTTGACGCTGTGTGTGTGCTGCTGTATACCAACAAATGAGTGAGACACTTCTGTGAGTTG	482			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					
554	GTCGCGCGCGGATGTGGGGCTTTTCAACAAGACAGACAGATGCCAAGATGTGAAGCAG	613			
483	CTGTAGTGCCTGCATCAAGA-----AGATTATGTTTCCAGAGAAGACTTCACTCAA	536			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					
614	TTCTATGACACAGGCGCTTACAGCAGCCGTGTGGATGATGAGCCCAACACCCCAAGGCT	673			
537	GTTGTGAACACACATGAAGAGGCTCAAGTGTCTGTGCTTACCAACTATACGATTTT	596			
111 111 111 111 111 111 111 111 111 111 111 111 111 111					

DB	674	GTGGGAGACCTTCAGACAGACGCTTACCTGCTGTGGCTCCAGACACTGACTCTTG	733
OY	597	GAGGACTCA	605
DB	734	ACCACCTCA	742
RESULT	8		
MUSTALLA			
LOCUS	MUSTALLA	1720 bp	mRNA
DEFINITION	Mouse mRNA for PE31/TALLA, complete cds.		ROD
ACCESSION	D26483		06-FEB-1999
NID	9685220		
VERSION	D26483.1	GI:685220	
KEYWORDS	PE31/TALLA; T-ALL associated antigen A15 counterpart; CCG-B7 counterpart.		
SOURCE	Mus musculus (strain:Balb/c) adult 8w whole brain cDNA to mRNA, clone_lib:library of M.Nagira clone:MPE31-I.		
ORGANISM	Mus musculus		
REFERENCE	Euhayota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
AUTHORS	Rodentia; Sciurognathii; Muridae; Murinae; Mus.		
TITLE	1 (bases 1 to 1720)		
JOURNAL	Isihkawa, I.		
REFERENCE	Submitted (11-JUN-1994) to the DDBJ/EMBL/GenBank databases. Izumi		
AUTHORS	Isihkawa, Shionogi Institute for Medical Science; 2-5-1 Mishima,		
TITLE	Settsu, Osaka 566, Japan (Tel:06-382-2612(ex.478), Fax:06-382-2598)		
JOURNAL	2 (bases 1 to 1720)		
FEATURES	Nagira, M., Isihkawa, I., Fujikawa, K., Takagi, S. and Yoshie, O.		
source	Molecular Cloning and Expression of mouse PE31 (TALLA) Unpublished (1994)		
	Location/Qualifiers		
	1..1720		
	/organism="Mus musculus"		
	/strain="Balb/c"		
	/db_xref="taxon:10090"		
	/clone="MPE31-I"		
	/clone_lib="library of M.Nagira"		
	/dev_stage="adult 8w"		
	/tissue_type="whole brain"		
	1..5		
	6..740		
	/note="murine counterpart of A15 (T-ALL associated antigen) and CCG-B7 which is expressed in the brain and is containing CCG triplet repeats"		
	/codon_start=1		
	/product="murine PE31 (TALLA)"		
	/protein_id="BA05493.1"		
	/db_xref="PID:d1006037"		
	/db_xref="PID:g685221"		
	/db_xref="GI:685221"		
	/translation="METRPVITCLTLTIYSFEVITLGVILLAVGWKGLTGTIYISLIANSNAPVLVIGCTTIVPELFCGPAICRSPMMKIKYAFSLVFLAEIVAGIISGPFREIKDITPLRTYTDANQNNNGNDESRANDYHQRSLSCGCVONYTWSSPFPELDHIFPSCCENETDCNPDLIDHINIVATKYNQCYDVLVSFMEINGIAGVAFGIFAFSOLIGMLACCCSRITINQYEMV"		
	6..737		
	/product="murine PE31/TALLA"		
	740..1719		
	polya_signal	1698..1703	
	polya_site	1720	
	BASE COUNT	400 a 428 c 406 g 486 t	
	ORIGIN		
Query Match	5.6%:	Score 71.6:	DB 12: Length 1720:
Best Local Similarity	46.2%:	Pred. No. 3.8e-10:	
Matches 331: Conservative	0:	Mismatches 364:	Indels 21: Gaps 2:
OY	120	ACCAACCTGTGATTAAGCAACGATGATCTTCACATTTGCTATCTTT	179
DB	12	ACCAACCTGTGATTAAGCAACGATGATCTTCACATTTGCTATCTTTG	71

OY		180	CTGTGGGACGCCCTTGTCAGTGGGCATCTGGGTGTCATCATGATGGGGCATTCTT	239
Dd		72	ATCACTGGGGGTATCTCTGTGTGGCCCTTTGGAGCTCTGGGAACACTACTTTGGGAACCTAT	131
OY		240	CTGAAGATCTTGGGGCACCTGTCTGCACAGTGCATGCAGTTGTGTACAGGTGGGCTACTTC	299
Dd		132	-----ATCTCCGTGATGTGTGAAGAACTCCACAAGAATGCTCCCATGTG	173
OY		300	CTCATCGCAGCCGGCGTTGTGTCTTTTGTCTTTGGTTTCTTGCGCTGTANGTGTCTAAG	359
Dd		174	CTCATTTGGAACCGGCACACCACCATCTGTGTGTGTGGCCCTTTGTGATGTGCTCTACATGC	233
OY		360	ACTGAGACAGATGTGGCCCTGTGCAGTTCTCTTCATCTCTCTCATCTTAATTTCT	419
Dd		234	CCTGTGTAGTCATGATGTGTGAAGTGTATGTGCATGTCTGTCTCTCTGTGTCTGTGT	293
OY		420	GAGTTTGACGCTGTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	479
Dd		294	GAGCTTGT	353
OY		480	TTCGT	539
Dd		354	AGGACTTACACAGGATGTGCATGAGAACTACATG---CAATGATGAGAGAGCGGGGCT	410
OY		540	TGGACACACCATGATAAGAGGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	599
Dd		411	GTGGACCATGTGTGCACGCGCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	470
OY		600	GACTCACCTTACTTCAAAGAGAACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	659
Dd		471	AGCAGCCCCCTACTTCTGT	530
OY		660	ACCAACACAGCCCAATGAAAACCTGTGCACCAAGCAAAAGGCTCACGACCAAAAGTAGAGGT	719
Dd		531	TGCATCTCCCCTGT	590
OY		720	TGCTTCATCAGCTTTTGTATGACATCCGAACTATGACGTACCGTGGTGTGTGTGTGTGTGT	779
Dd		591	TGTTTGT	650
OY		780	GCTGGAATTTGGGGGCTGT	835
Dd		651	TTTGGATTTGCTTCTTCCAGTTGATGTGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	706
RESULT		9		
LOCUS		AF116600	711 bp mRNA	PRI 01-FEB-1999
DEFINITION		Pan troglodytes CD81 mRNA, complete cds.		
ACCESSION		AF116600		
NID		94206172		
VERSION		AF116600.1 GI:4206172		
KEYWORDS		.		
SOURCE		chimpanzee.		
ORGANISM		Pan troglodytes		
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS		Eutheria; Primates; Catarrhini; Hominoidea; Pan.		
JOURNAL		Levy, S. and Kuo, C.C.		
REFERENCE		CD81 cDNA sequence derived from chimpanzee cells		
AUTHORS		Unpublished		
TITLE		2 (bases 1 to 711)		
JOURNAL		Levy, S. and Kuo, C.C.		
REFERENCE		Direct Submission		
AUTHORS		Submitted (24-Dec-1998) Medicine/Oncology, Stanford, Stanford, CA		
TITLE		94305, USA		
JOURNAL		Location/Qualifiers		
FEATURES		1..711		
SOURCE		/organism="Pan troglodytes"		
		/db_xref="taxon:9598"		
		/cell_line="Jenice"		
		/note="EBV-transformed cells; amplified using Homo sapiens		
		CD81-derived primers"		

primer_bind	<1..14
CDS	1..711
	/codon_start=1
	/product="CD81"
	/protein_id="AA011440.1"
	/db_xref="PID:g4206173"
	/db_xref="GI:4206173"
	/translation="MEVEGCTCKIKYILFYPNFVYLAGVILGVALMLRHPDTNL LYLEAGKRPAPNFTYVIGVILYILAVGAAMTVVGLGCGAIOESCLLGTFTLVILF ACEVLAEGPVPNFDQJAKDKQKQFYDADLOOAAYDDANNAAVKTGHEFTLDCGSS TLPLATTVSLKNNKCPGSGNSINSLMFREDCHOKIDLFSEKLYLIGIAIVAVIMIF EMLSWVLCGGIRNSVY"
BASE COUNT	136 a 212 c 200 g 163 t
ORIGIN	complement(693..>711)
Query Match	5.5% Score 70.6; DB 11; Length 711;
Best Local Similarity	49.7% Pred. No. 6.4e-10;
Matches 243; Conservative	0; Mismatched 234; Indels 12; Gaps 2;
QY	129 TGCCTCAGCTTCATTAAAGCCATGATGATCCCTTCATATTGCTCATCTTCTGTGTGT 188
Db	16 TGCACCAAGTGCATCAATACCTGCTCTCTCTTCATTTTCGCTTCTGTGGCTGTGGA 75
QY	189 GCAGCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGG-----GCATCTTTTG 242
Db	76 GCGCGATCTCTGGGAGTGGCCCTGGGCTCCGCATGACCCGACACCAACCTCTG 135
QY	243 AAGATCTTGGGGCCACTCTCTCCAGTGCATGCAAGTTGTTCACAGTGGGCTACTCTC 302
Db	136 TATCTGAGCTGGGGAGACAAAGCCCGCCCAACACTTCTATGATGAGCATCTACATCTC 195
QY	303 ATCCGAGCCGGCGTGTGTGTCTTGTCTTGTGTCTTGTCTGTGTGTGTGTGTGTGT 362
Db	196 ATCCGTGTGGGCGCTGTATGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 255
QY	353 GAGAGCAAGTGTGCCCTCTGTACGCTTCTTCTTCATCTCTCTCATCTTCTCATTTGCTGAG 422
Db	256 GAATCCAGT 315
QY	423 GTTGCACTGCTGT 482
Db	316 GTGGCGCGGCGCATCTGGGGCTTTGTTCACAAAGGACCCAGATGCCACAGAGTGTGAGCG 375
QY	483 CTGTAGT 536
Db	376 TTTCTGACCAAGGCGCTTACACAGAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 435
QY	537 GTGTGAGCAACACCATGAAAGGCGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 596
Db	436 GTGTGAGAGACCTTCCAGAGACGTTTACAGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 495
QY	597 GAGGACTCA 605
Db	496 ACCACCTCA 504
RESULT 10	
AF116599	
LOCUS	AF116599 711 bp mRNA PRI 01-FEB-1999
DEFINITION	Chlorocebus aethiops CD81 mRNA, complete cds.
ACCESSION	AF116599
NTD	g4206170
VERSION	AF116599.1 GI:4206170
KEYWORDS	
SOURCE	African green monkey.
ORGANISM	Chlorocebus aethiops
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
	Chlorocebus.
REFERENCE	1 (bases 1 to 711)
AUTHORS	Levy, S. and Kuo, C.C.

TITLE African green monkey CD81 cDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 711)  
AUTHORS Levy, S. and Kuo, C.C.  
TITLE Direct Submission  
JOURNAL Submitted (24-DEC-1998) Medicine/Oncology, Stanford, Stanford, CA 94305, USA

FEATURES  
source Location/Qualifiers  
1..711  
/organism="Chlorocebus aethiops"  
/db\_xref="taxon:9534"  
/cell\_line="COS"  
/note="amplified using Homo sapiens CD81-derived primers"  
1..14  
primer\_bind  
CDS  
1..711  
/codon\_start=1  
/product="CD81"  
/protein\_id="AA011439.1"  
/db\_xref="PID:94206171"  
/db\_xref="GI:4206171"  
/translation="MGVEGCTCICRYLLEFNFVFMVAGVTLGVALMRHPORTNL  
LYLEGGDPAPWTFYVGYIILANGANMFVFLCYGAIOSQCLSTFPCYILF  
ACEVAGINGFVNDQIADVAQFYDQADQAVDDANNAKAVKTHETLDCGSS  
TLAALTSTVLKNNLCPSGSNIISNLKDKHOKIDELFSGLYLIGIAIVAVYIMIF  
EMLSMVLCCGIRNSVY"  
complement(693..>711)  
BASE COUNT 140 a 202 c 200 g 169 t  
ORIGIN

Query Match 5.3%; Score 67.8; DB 11; Length 711;  
Best Local Similarity 49.9%; Pred. No. 4.2e-09;  
Matches 233; Conservative 0; Mismatches 222; Indels 12; Gaps 2;

QY 128 GTGCTTACCTTATTAAGACATGATGATCCTCTTCAATTTGCTATCTTTGTGTGG 187  
DB 15 GTGTACCAAGTGCATAGATACCTGCTCTTCCTCAATTGCTTCTGGCTGGCTGG 74  
QY 188 TGCAGCCCTGTGGGAGTGGGATCTGGGTGTGTAATGAGG-----GCATCCTTCT 241  
DB 75 AGCGGTGATCTTGAGTGTAGCTTGTGGCTTCCGCCATGAGCCGACATCACTACCTCT 134  
QY 242 GAAGATCTTGGGCACTGTCTGCAAGTGCATGAGATTTGCAAGTGGGCTACTTCTCT 301  
DB 135 GTATCTGAGCTGGAGACAAGCCCTGACCCATACCTTCTAGTGGCATCTACATCTCT 194  
QY 302 CATCGACCGGGGTGTGTCTTCTGCTTCTGGCTTCTGGGCTGTATGTTGTTAGAC 361  
DB 195 CATGGCCGTGGGGCTGTGATGATGCTTGGCTTCTGGGCTGTGAGGGGCTATGCA 254  
QY 362 TGAAGACAGTGTGCGCTGTGATCTTCTCATCTCCCTCATCTCATCTTCTGATG 421  
DB 255 GGAATCCCAAGTGTCTGTGGGAGCTTCTTCACTGTCTGTATCTCTGTTTCCCTG 314  
QY 422 GGTGAGCTGTCTGTGTGCTTGTGTGATGACACCAATGCTGAGACACTTCTGAGCTT 481  
DB 315 AGTGGAGCGGGCATCTGGGGCTTTGTCAACAGACAGCAAGATTGCCAAGATGAGAGA 374  
QY 482 GCTGTGATGCTTCCCATACAGAAATTAATGTTCCAGGAAGACTTACT-----CA 535  
DB 375 GTTCTATGACAGAGCCCTTACAGAGGCTGTGGTGTGAGAGGCCCAATGCCAAGC 434  
QY 536 AGTGTGAGACACCATGAAGAGGCTCAAGTGTGTTGGTTTTCAGCA 582  
DB 435 CGTGTGAGAACCTTCCACGAGACGCTTGTGCTGTGCTCCAGCA 481

RESULT 11  
HUMCD53 1452 bp mRNA PRI 01-NOV-1994  
LOCUS HUMCD53 1452 bp mRNA PRI 01-NOV-1994  
DEFINITION Human cell surface antigen (CD53) mRNA, complete cds.  
ACCESSION M60871  
NID 9160140

VERSION M60871.1 GI:180140  
KEYWORDS cell surface antigen; type III integral membrane protein.  
SOURCE Human promyelocytic tumor cell line HL60, cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1452)  
AUTHORS Amiot, M.  
TITLE Identification and analysis of cDNA clones encoding CD53. A  
pan-leukocyte antigen related to membrane transport proteins  
JOURNAL U. Immunol. 145 (12), 4322-4325 (1990)  
MEDLINE 91079522  
FEATURES  
source Location/Qualifiers  
1..1452  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="promyelocytic tumor cell line HL60"  
/map="unassigned"  
1..1452  
/gene="CD53"  
/note="G00-127-521"  
1..1452  
/gene="CD53"  
74..733  
/gene="CD53"  
/codon\_start=1  
/db\_xref="GDB:G00-127-521"  
/product="cell surface antigen"  
/protein\_id="AA01951.1"  
/db\_xref="PID:9180141"  
/db\_xref="GI:180141"  
/translation="MGMSSTKLKLYLEFNFLLPMIGCCILGFGIYLLHNFGLV  
HNLSLTGAVFVIGSIIVAVFLGMSIKKKELMSFILLIILAEVTLAL  
LFVYRKLNEVYAGLDSIRYSDSTRAMSTISFLOCCTNSDPTSGPPAS  
CPSRKVEGCTAKRLMFHSNFLTIGLITICVIEVLSFALTLQIDKISQTLG  
L"  
BASE COUNT 373 a 342 c 314 g 423 t  
ORIGIN

Query Match 4.9%; Score 62.8; DB 9; Length 1452;  
Best Local Similarity 51.2%; Pred. No. 1.3e-07;  
Matches 173; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

QY 257 ACTGTGTCATGTCAGATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTC 316  
DB 196 AGTGTCTTCCATTAACCTCCCTCCCTCAGCTGGGCAATGTGTTGTATCAGTGGCTC 255  
QY 317 TGTGCTTGTGCTTGTGTTTCTGGGCTGTATGTCATGTCATGTCATGTCATGTCATG 376  
DB 256 TATTATCATGTAAGTGGCTTCTGGGCTGTATGTCATGTCATGTCATGTCATGTCATG 315  
QY 377 CCTGTGACCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 436  
DB 316 GCTTATGTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 375  
QY 437 GGTGCGCTGTGTACACACAATGGCTGAGACATCTTCTGACCTGTGCTGTAGTGGCTGC 496  
DB 376 CCTGCTCTTGTATATGACACAGAACCTGAATGATGATGGC--TAAGGCTTACCGA 432  
QY 497 CATCAAGAAAGTATATGTTGTTCCAGGAAGACTTCTCAAGTGTGGAACACCATGAA 556  
DB 433 CAGCATCCACCTTACACTACAGACAAATAGACCAAGCAGAGGTGGACTCCATCAGTC 492  
QY 557 AGGCTCAAGTGTGTGCTTCCACCACTATACGATT 594  
DB 493 ATTCTGACAGTGTGTGATTAATGACAGAGTATT 530

RESULT 12  
HUMCD53GLY 1480 bp mRNA PRI 03-MAR-1994  
LOCUS HUMCD53GLY 1480 bp mRNA PRI 03-MAR-1994  
DEFINITION Human CD53 glycoprotein mRNA, complete cds.

FEATURES	SOURCE
CD5	
polya_signal	1456, 1461
BASE COUNT	378 a 353 c 318 g 431 t
ORIGIN	

M37033  
 M180142  
 M37033.1 GI:180142  
 CD53 antigen.  
 Human lymphocyte LAK-cell, cDNA to mRNA, clone CD53-65.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 1480)  
 Angellsova, P., Vlcek, C., Stefanova, I., Lipoldova, M. and Horejsi, V.  
 The human leucocyte surface antigen CD53 is a protein structurally  
 similar to the CD37 and MHC CX-44 antigens  
 Immunogenetics 32, 281-285 (1990)  
 91055810  
 Draft entry and computer-readable sequence for [Immunogenetics  
 (1990) In press] kindly submitted  
 by V.Horejsi, 25-JUL-1990.  
 Location/Qualifiers  
 1..1480  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 94..753  
 /codon\_start=1  
 /product="CD53 glycoprotein"  
 /protein\_id="AAA35663.1"  
 /db\_xref="GI:9180143"  
 /db\_xref="GI:180143"  
 /translation="MGMSLKLLKLVLPFENLFWICGCCIGFGIYLLIHNNFCVLF  
 HNPDSLTGNEFVIVGSIIMVAVFLGMSGIKENKCLMSEFILLIILAEVTLAIL  
 LVEYQRLNEVAVGLDSDIRHYSDSSTKRAMSDISFLDCCGINGSDWTS6PPAS  
 CPSDRKVEGCYAKARLWFHNSNFIYIGITICVIEVLGMSFALTLCQIDKITSOTIG  
 L"

Query Match	Score 62.8	DB 9	Length 1480
Best Local Similarity	51.28	Pred. NO.1.3e-07	
Matches 173	Conservative 0	Mismatches 162	Indels 3
			Gaps 1
QY 257	ACTGCTCCAGTGCATGACATTTGTGTCACAGTGGGCTACTCTCATCGACGGCGGT	316	
Db 216	AGTGTCTTCATACACCTCCCTCCCTCAGGTGGGCATGTGTTCATCGTGGGCTC	275	
QY 317	TGTGCTCTTTGCTCTTGGTTCTCTGGGCTGCTATGCTGAAGACTGAAGCAAGTGTGC	376	
Db 276	TATATCATGTGAGTGGCTCTCTGGGCTGCTATGGGCTCTATCAAGGAAACAACTGTCT	335	
QY 377	CCTGTGACGTTCTCTTCATCCTCCTCCTCATCTTCATGGTGAAGTTGACAGTGTGT	436	
Db 336	GCTTATGTCTCTCTTCATCTCTGCTCTGCTATTTCTCTCTGCTGAAGTACCTTGGCCAT	395	
QY 437	GGTGCCTTGTGTGACACCAACAATGGCTGAGACACTCTGACGTTGCTGTGATGCTTGC	496	
Db 396	CCTGCTCTTGTATATGAAACAAGAACTGAATGATATGGC---TAAAGGCTGACCGA	452	
QY 497	CATCAAGAAAGTTATGTGTTCCCAAGAACTTCACTCAAGTGTGGAAACACACATGAA	556	
Db 453	CAGCATCCACGGTTACCACTAGAACAAAGACCAAGGACAGGTGGGACTCATCTCAAGTC	512	
QY 557	AGGCTCAAGTGTGCTGGCTTCCACCAACTATACGAGATT	594	
Db 513	ATTTCGACGTGTGTGGTATTAATGGCAGCAGTATTT	550	
RESULT 13			
LOCUS	RNU19894		
DEFINITION	RNU19894	1303 bp	mRNA
ACCESSION	U19894		ROD
	g1142641		25-MAR-1997
			complete cds.
			antibody mRNA,

VERSION	U19894.1	GI:1142641
KEYWORD		
SOURCE	Norway rat.	
ORGANISM	Rattus norvegicus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS	1 (bases 1 to 1303)	
TITLE	Gelsert,E.E. Jr., Murphy,T.P., Irwin,M.H. and Larjava,H.	
JOURNAL	A novel cell adhesion molecule, G-CAM, found on cultured rat glia	
MEDLINE	Neurosci. Lett. 133 (2), 262-266 (1991)	
REFERENCE	2 (bases 1 to 1303)	
AUTHORS	Irwin,M.H. and Gelsert,E.E. Jr.	
TITLE	The upregulation of a glial cell surface antigen at the astrocytic	
JOURNAL	scar in the rat	
MEDLINE	Neurosci. Lett. 154 (1-2), 57-60 (1993)	
REFERENCE	3 (bases 1 to 1303)	
AUTHORS	Gelsert,E.E. Jr., Yang,L. and Irwin,M.H.	
TITLE	Astrocyte growth, reactivity, and the target of the	
JOURNAL	antiproliferative antibody, TAPA	
MEDLINE	J. Neurosci. 16 (17), 5478-5487 (1996)	
REFERENCE	4 (bases 1 to 1303)	
AUTHORS	Gelsert,E.E.	
TITLE	Direct Submission	
JOURNAL	Submitted (17-JAN-1995) Eldon E. Gelsert Jr., Department of Anatomy	
FEATURES	and Neurobiology, University of Tennessee, Memphis, 855 Monroe	
SOURCE	Ave., Memphis, TN 38163, USA	
	Location/Qualifiers	
	1..1303	
	/organism="Rattus norvegicus"	
	/strain="Sprague-Dawley"	
	/db_xref="taxon:10116"	
	/clone="1.1, spl, sp2"	
	/clone_1b="astrocyte cDNA lambda gcl1 (Clontech No.	
	RL1042b)"	
	/sex="Male"	
	/cell_type="astrocyte (Type 1 and Type 2)"	
	/tissue_type="cerebral cortex"	
	1..222	
	223..933	
	/note="TAPA"	
	/codon_start=1	
	/product="target of the antiproliferative antibody"	
	/protein_id="A053103.1"	
	/db_xref="PIR:G1142642"	
	/db_xref="GI:1142642"	
	/translation="MGVGGCTKCIKYLLFVFNFVNLAVGVIIGVALMLRHPDPTLL	
	LYLEADGKPAFTFYVGIYLLIIVAGVMVVGAGCYGAIOESCILGTFETCLVILF	
	ACEVAAGIMGFVNKDIAKVDKQFYDQALQAAVMDPANAKAVKTFHETLNCGSN	
	TLTITITAVLRNSICPSSSNSTFQLLKEDHKIDELFSKIVYLIGIAIVAVIMIF	
	EMLSWTCGGINSSSY"	
	EMLSWTCGGINSSSY"	
BASE COUNT	3'UTR 934..1303	
ORIGIN	257 a 375 c 347 g 324 t	
Query Match	4.88; Score 62.2; DB 12; Length 1303;	
Best Local Similarity	53.58; Pred. No.1.9e-07;	
Matches 177; Conservative	0; Mismatches 148; Indels 6; Gaps 2;	
129	TGCTTCACCTCAATTAACACCATGATGATCTCTTCATATTGCATCATTTCTGTGTCGT	188
111	111 111 111 111 111 111 111 111 111 111 111 111 111 111 111	111
238	TGCACCAATATCATCAATATACCTGCTCTTCCTTCATATTCGTCCTTGTGGCTGGCTGA	297
189	GCAGCCCTGTGGCAGTGGGCATCTGGGTGTCAATGCATGCATGGGCGATCCTTCTGAAGAT	248
111	111 111 111 111 111 111 111 111 111 111 111 111 111 111	111
298	GGTGTGATCCATAGGTATGCTGTGTGTGGCATATATCCACAGACACACCTGTGTC	357
249	TTGGGGCCACGTC-----TCGTCCAGTGGCATGCAGTTTGTCACGTC-GGCTACTTCTC	302
111	111 111 111 111 111 111 111 111 111 111 111 111 111 111	111
358	TACCTGTGAACGTGGAGACAACACGACCTGACACCTTCATCTGTGTGGGCATTCATATCTC	417



```

/dp_xref="axon:9606"
/dev_string="85 years"
/sex="female"
/issue_type="frontal Cortex"
/issue_id="Stratagene lambda ZAP II"
BASE COUNT      434 a      447 c      399 g      512 t
ORIGIN

```

**This Page Blank (uspto)**





Db	361	CTGAGACGAAGTGTGCCCTGCTGACGTTCTTCTTCAACCTCCCTCATCTTCATTTCGTC	420
Qy	421	AGGTTGCAGCTGCTGTGTGTGGCTTGTGGTGTACACCAATGGCTGAGCACTTCTGACGT	480
Db	421	AGGTTGACGGCTGTGTGTGGCTTGTGGTGTACACCAATGGCTGAGCACTTCTGACGT	480
Qy	481	TGCTGTAGTGTCCCTGCCATGAGAAAGATTATGTGTTCCAGAGAGACTTCACTCAAGTGT	540
Db	481	TGCTGTAGTGTCCCTGCCATGAGAAAGATTATGTGTTCCAGAGAGACTTCACTCAAGTGT	540
Qy	541	GGAACACCAACCATTGAAGAGGCTCAAGTGTGTGGCTTCAACCAATATACGATTTTGAAG	600
Db	541	GGAACACCAACCATTGAAGAGGCTCAAGTGTGTGGCTTCAACCAATATACGATTTTGAAG	600
Qy	601	ACTACCCCTACTTCAAGAGACAGTGGCTTTCGCCATTCCTGTGCATGACACAGCA	660
Db	601	ACTACCCCTACTTCAAGAGACAGTGGCTTTCGCCATTCCTGTGCATGACACAGCA	660
Qy	661	CCAACACAGCCCAATGAACCTGCAACAGCAAAAGGCTCACGACCAAAAAGTAGAGGTT	720
Db	661	CCAACACAGCCCAATGAACCTGCAACAGCAAAAGGCTCACGACCAAAAAGTAGAGGTT	720
Qy	721	GCTTCAATCAAGCTTTGTATGACATCCGAATATACAGTCAACGCTGGGTGTGGCAG	780
Db	721	GCTTCAATCAAGCTTTGTATGACATCCGAATATACAGTCAACGCTGGGTGTGGCAG	780
Qy	781	CTGGAATTTGGGGGCTGAGACTGGCTGCCATGATTTGTTCATGTATCTGTCTGCATTC	840
Db	781	CTGGAATTTGGGGGCTGAGACTGGCTGCCATGATTTGTTCATGTATCTGTCTGCATTC	840
Qy	841	TACAATAGTCCACTTCTGCTCTGCCACTACACACACACATGGGAACCTGTGAAGAGGC	900
Db	841	TACAATAGTCCACTTCTGCTCTGCCACTACACACACACATGGGAACCTGTGAAGAGGC	900
Qy	901	ACCTGCGCAACACAGTGTATGGGGAGGGGACAGAGATCTAACAATGTCACTTGGGCCA	960
Db	901	ACCTGCGCAACACAGTGTATGGGGAGGGGACAGAGATCTAACAATGTCACTTGGGCCA	960
Qy	961	GAATGAGACCTGCCCCCTTCTCTCAGACTTGGGGGTAGATAGGAGACCACTCCTTTAGCG	1020
Db	961	GAATGAGACCTGCCCCCTTCTCTCAGACTTGGGGGTAGATAGGAGACCACTCCTTTAGCG	1020
Qy	1021	ATGCGTCACTTCCCTTCATTTGGTGGGTGATGGTGGGGGATTCAGAGCCTCTAAG	1080
Db	1021	ATGCGTCACTTCCCTTCATTTGGTGGGTGATGGTGGGGGATTCAGAGCCTCTAAG	1080
Qy	1081	GTACCAAGTTCTGTGTGCCAATTCGCCAGTCTATTAAACCTTTGATATGCCCCCTTAGGC	1140
Db	1081	GTACCAAGTTCTGTGTGCCAATTCGCCAGTCTATTAAACCTTTGATATGCCCCCTTAGGC	1140
Qy	1141	TAGTGTGATATCCCAAGTCTCTACTGAGGGAGTGAAGAAAGCAATTTATAGCTGGGCAT	1200
Db	1141	TAGTGTGATATCCCAAGTCTCTACTGAGGGAGTGAAGAAAGCAATTTATAGCTGGGCAT	1200
Qy	1201	AAGTGAATCAGCAGACCTCTGGTGGATGTGTAGAAAGCACTTCAAAATGCATTAACC	1260
Db	1201	AAGTGAATCAGCAGACCTCTGGTGGATGTGTAGAAAGCACTTCAAAATGCATTAACC	1260
Qy	1261	TGTTACAATGTTAAAAAATTTTTTTTTT 1289	
Db	1261	TGTTACAATGTTAAAAAATTTTTTTTTT 1289	

RESULT	2
ID	V61202
AC	V61202 standard; CDNA: 1289 BP.
DT	06-JAN-1999 (first entry)
DE	Full length cDNA sequence of prostate tumour clone N1-1862
KW	Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS	Human sapiens.
PN	MO9837093-A2.

PD 27-AUG-1998 .  
 PE 25-FEB-1998; U03492.  
 PR 09-FEB-1998; US-020956.  
 PR 25-FEB-1997; US-806099.  
 PR 01-AUG-1997; US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI; 98-609886/51.  
 P-PEDB; W71870.  
 PR Polypeptides comprising immunogenic portions of prostate proteins -  
 PR used in a vaccine for the treatment of prostate cancer  
 PS Claim 3; Page 80; 10pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
 CC was identified by analysis of a subtracted cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SQ Sequence 1289 bp; 292 A; 335 C; 328 G; 334 T;

Query Match	100.0%	Score 1289;	DB 1;	Length 1289;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1289;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	AGCCAGGGGCTCCCTTGCTGGCTGGCCACTGAGTGGCAACCGGGGAGCTGTTTGTGCTTT	60	
DB	1	AGCCAGGGGCTCCCTTGCTGGCTGGCCACTGAGTGGCAACCGGGGAGCTGTTTGTGCTTT	60	
QY	61	GTGGAGCTCAGCAAGTTCCCTCTTTCAAGACTCACTGCCCAAGAGCCCTGAAACAGGACCA	120	
DB	61	GTGGAGCTCAGCAAGTTCCCTCTTTCAAGACTCACTGCCCAAGAGCCCTGAAACAGGACCA	120	
QY	121	CCATCAGAGCTTCAGTCTCATTTAAGACCATGATATCTCTCAATTGGCATCTTC	180	
DB	121	CCATCAGAGCTTCAGTCTCATTTAAGACCATGATATCTCTCAATTGGCATCTTC	180	
QY	181	TGTGTGTGTCAGACCCCTGTGGCAGTGGGCATCTGGGTTCATCGATGGGCATCTTTC	240	
DB	181	TGTGTGTGTCAGACCCCTGTGGCAGTGGGCATCTGGGTTCATCGATGGGCATCTTTC	240	
QY	241	TGAGATCTTGGGGCCACTGTGTGTCAGTGCATGCCATGCAAGTTGTCAAGTGGGCTACTTC	300	
DB	241	TGAGATCTTGGGGCCACTGTGTGTCAGTGCATGCCATGCAAGTTGTCAAGTGGGCTACTTC	300	
QY	301	TGATGGCAGCGGGCTGTGGCTTTGGCTTTGGCTTCTGGCGCATGGTGTAGTA	360	
DB	301	TGATGGCAGCGGGCTGTGGCTTTGGCTTTGGCTTCTGGCGCATGGTGTAGTA	360	
QY	361	CTGAGAGCAAGTGTGCCCTGTGAGCTCTTTCATCCTCCTCATCTTCATTTGCTG	420	
DB	361	CTGAGAGCAAGTGTGCCCTGTGAGCTCTTTCATCCTCCTCCTCATCTTCATTTGCTG	420	
QY	421	AGGTTGCAGCTGCTGTGTCGCTTGTGTGACACCAATGGCTGAGCATCTTCTGACGT	480	
DB	421	AGGTTGCAGCTGCTGTGTCGCTTGTGTGACACCAATGGCTGAGCATCTTCTGACGT	480	
QY	481	TGCTGTGTGTGCTGCTGCCATCAAGAAAGTTATGGTTCACAGGAAGACTTCACTCAAGGT	540	
DB	481	TGCTGTGTGTGCTGCTGCCATCAAGAAAGTTATGGTTCACAGGAAGACTTCACTCAAGGT	540	
QY	541	GGAAACCAACCATGAAAGGGCTCAAGTCTGTGCTTACCAACATATACGATTTTGGAG	600	
DB	541	GGAAACCAACCATGAAAGGGCTCAAGTCTGTGCTTACCAACATATACGATTTTGGAG	600	
QY	601	ACTCACCTTACTTCAAGAGAACAGTGCCTTCCCATCTGTGTGCAATGACCAACGTCA	660	
DB	601	ACTCACCTTACTTCAAGAGAACAGTGCCTTCCCATCTGTGTGCAATGACCAACGTCA	660	
QY	661	CCAAACAGCCCATGAAACCTGTGACCAAGCAAAAGGCTCAGAGCCAAATAAGAGGTT	720	
DB	661	CCAAACAGCCCATGAAACCTGTGACCAAGCAAAAGGCTCAGAGCCAAATAAGAGGTT	720	



PR 22-AUG-1997: US-056908.  
 PR 22-AUG-1997: US-056909.  
 PR 22-AUG-1997: US-056910.  
 PR 22-AUG-1997: US-056911.  
 PR 05-SEP-1997: US-057650.  
 PR 05-SEP-1997: US-057651.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Bednarik DP, Brewer LA, Carter KC, Duan R, Edner R, Endress GA,  
 PI Feng P, Ferris AM, Fischer CL, Graves KA, Greene JM, Hu JS,  
 PI Kyaw H, Lafleur DR, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA,  
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;  
 DR WPI: 98-60987/51.  
 DR P-PSDB: W75060.  
 PT New isolated human genes and the secreted polypeptides they encode  
 PT - useful for diagnosis and treatment of e.g. cancers, neurological  
 PT disorders, immune diseases, inflammation or blood disorders  
 PS Claim 1: Page 167; 447pp: English.  
 CC This sequence represents a nucleic acid molecule which encodes a secreted  
 CC human protein. The gene number, and the clone it is derived from, are  
 CC detailed in the descriptor line. The gene can be used to generate fusion  
 CC proteins by linking to the gene to a human immunoglobulin Fc portion  
 CC (e.g. V4154) for increasing the stability of the fused protein as  
 CC compared to the human protein only.  
 CC The invention relates to 70 novel genes and their fragments (nucleic acid  
 CC sequences: V34154-V34276; amino acid sequences W75057-W75179) which  
 CC are useful for preventing, treating or ameliorating medical conditions  
 CC e.g. by protein or gene therapy. Also, pathological conditions can be  
 CC diagnosed by determining the amount of the new polypeptides in a sample  
 CC or by determining the presence of mutations in the new polynucleotides.  
 CC Specific issues are described for each of the 70 polynucleotides, based on  
 CC which tissues they are most highly expressed in (see V34154 for described  
 CC uses).  
 SO Sequence 1376 BP; 316 A; 354 C; 356 G; 347 T;

Query Match 99.1%; Score 1277.6; DB 1; Length 1376;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1288; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

QY 1 AGCCAGGCGGCTGCTGCGCCGCTCCACTCATGAGCAACCCGGGACTGTTTGTCTTT 60  
 DB 53 AGCCAGGCGGCTGCTGCGCCGCTCCACTCATGAGCAACCCGGGACTGTTTGTCTTT 112  
 QY 61 GTGAGACCTTCAGAGTTCCTTTTGAAGTCACTGCAAGAGCCCTGAACAGAGCA 120  
 DB 113 GTGAGACCTTCAGAGTTCCTTTTGAAGTCACTGCAAGAGCCCTGAACAGAGCA 172  
 QY 121 CCATGAGTGTCTCAGCTTCATTAAAGACATGATGATCTTCAATTGCTCATCTTTC 180  
 DB 173 CCATGAGTGTCTCAGCTTCATTAAAGACATGATGATGATCTTCAATTGCTCATCTTTC 232  
 QY 181 TGTGTGTGTGAGCCCTGTGGGAGTGGGCAATGCGGTCAATGATGGGCAATCTTTC 240  
 DB 233 TGTGTGTGTGAGCCCTGTGGGAGTGGGCAATGCGGTCAATGATGGGCAATCTTTC 292  
 QY 241 TGAAGATCTTCGCGGCACTGCTCCAGTCCATGCAATGCAATGCTGGCTACTTTC 300  
 DB 293 TGAAGATCTTCGCGGCACTGCTCCAGTCCATGCAATGCAATGCTGGCTACTTTC 352  
 QY 301 TCATGCGAGCGGCGGTGTGCTTGTCTTGTGTTCTTGGGCTGCTATGTGCTAAGA 360  
 DB 353 TCATGCGAGCGGCGGTGTGCTTGTCTTGTGTTCTTGGGCTGCTATGTGCTAAGA 412  
 QY 361 CTGAGAGCAAGTGTGCTGCTGAGCTTCTTCAATCCCTCCCTCATCTTATGCTG 420  
 DB 413 CTGAGAGCAAGTGTGCTGCTGAGCTTCTTCAATCCCTCCCTCATCTTATGCTG 472  
 QY 421 AGGTTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480  
 DB 473 AGGTTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 532  
 QY 481 TGCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540

DB 533 TGCTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 592  
 QY 541 GGAACACCCACATGAAAGGCTCAAGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 DB 593 GGAACACCCACATGAAAGGCTCAAGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 652  
 QY 601 ACTCACCTTACTTCAAGAGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660  
 DB 653 ACTCACCTTACTTCAAGAGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712  
 QY 661 CCAACACAGCCATGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 720  
 DB 713 CCAACACAGCCATGAAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 772  
 QY 721 GCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 780  
 DB 773 GCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 832  
 QY 781 CTGGAATTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
 DB 833 CTGGAATTGGGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 892  
 QY 841 TACAATTAAGTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
 DB 893 TACAATTAAGTCCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 952  
 QY 901 ACCCTGGCAAGAGAGATGATGGGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
 DB 953 ACCCTGGCAAGAGAGATGATGGGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1012  
 QY 961 GATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1019  
 DB 1013 GATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1072  
 QY 1020 GATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1079  
 DB 1073 GATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1132  
 QY 1080 GATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1139  
 DB 1133 GATGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1192  
 QY 1140 CTAGTGTGATCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1199  
 DB 1193 CTAGTGTGATCCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1252  
 QY 1200 TAAGTGAATTCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1259  
 DB 1253 TAAGTGAATTCAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1312  
 QY 1260 CTGTTACAAATGTTAAAAA 1289  
 DB 1313 CTGTTACAAATGTTAAAAA 1342

RESULT 4  
 V30916  
 ID V30916 standard; cDNA; 1605 BP.  
 AC V30916:  
 DT 14-SEP-1998 (first entry)  
 DE Human secreted protein AR415\_4 cDNA.  
 KW AR415.4; secreted protein; protein factor; human; ds.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 437..1162  
 FT sig\_peptide 476..514  
 FT mat\_peptide 515..1159  
 FT /note= "putative leader/signal peptide"  
 FT /tag= b  
 FT /tag= c  
 PD W09817687-A2.  
 PN 30-APR-1998.

PF 24-OCT-1997: U19590.  
 PR 24-OCT-1997: US-740274.  
 PR 25-OCT-1996: US-740274.  
 PA (GEM) GENETICS INST INC.  
 PA Agostino MJ, Jacobs K, Lavallie ER, McCoym JM, Metberg D,  
 PI Racine LA, Spaulding V, Treacy M,  
 PI WPI: 98-261426/23.  
 DR P-PDSB: W58380.  
 PT Nucleic acid encoding secreted protein from human cells - useful,  
 PT e.g. as immunomodulator, antitumour agent, promoters of tissue  
 PT growth, haemostatic and thrombolytic agents etc.  
 PS Claim 1: Page 66-67; 114pp: English.  
 CC This cDNA clone, designated AT415.4, codes for a novel human  
 CC secreted protein (see W58380). It was isolated from a human adult  
 CC retina cDNA library using methods selective for cDNAs that encode  
 CC secreted proteins. The clone is deposited in composite clone  
 CC ATCC 98323: an oligonucleotide (see V30933) is designed to isolate  
 CC the clone from the composite. The predicted AT415.4 amino acid  
 CC sequence shows homology to human M35252 and Co-029 tumour associated  
 CC antigens. Novel cDNA clones (see V30935-32) coding for human  
 CC secreted proteins (see W58580-90) are claimed. These can be used  
 CC for recombinant production of the secreted proteins for analysis,  
 CC characterisation, diagnostic or therapeutic use. They can also be  
 CC used as tissue or mol.wt. markers for chromosome identification,  
 CC to identify genetic disorders, to isolate new related DNA, as  
 CC sources of primers for PCR, to generate antibodies, and in  
 CC interaction trap assays. The secreted proteins may also have many  
 CC biological activities, e.g. cytokine, immunomodulator,  
 CC haematopoiesis regulating activity, tissue growth activity, activin  
 CC or inhibin activity, chemotactic or chemokinetic activity,  
 CC haemostatic and thrombolytic activity, receptor/ligand activity,  
 CC antiinflammatory, cadherin and tumour invasion suppressor activity,  
 CC and tumour inhibition activity. The proteins can be expressed in  
 CC vivo from DNA, introduced in gene therapy vectors.  
 Sequence 1605 BP: 360 A; 427 C; 399 G; 419 T;

Query Match	98.98;	Score 1274.8;	DB 1;	Length 1605;
Best Local Similarity	99.88;	Pred. No. 0;		
Matches 1287; Conservative	0;	Mismatches 2;	Indels 1;	Gaps 1;

OY	1	AGCCAGGCGTCCCTCTGCTGCTGCCACATCAGGCGCAACCCGGAGCTGTTTTGCTTT	60
Db	315	AGCCAGGCGTCCCTCTGCTGCTGCCACATCAGGCGCAACCCGGAGCTGTTTTGCTTT	374
OY	61	GTGAGGCGCTCAGCAGTTCCTCTTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA	120
Db	375	GTGAGGCGCTCAGCAGTTCCTCTTTTCAGAACTCACTGCCAAGAGCCCTGAACAGGAGCCA	434
OY	121	CCATGCAAGTGGTTAGCTTATTAAAGCAATGATGATCCCTTCAATTGCTCATCTTTC	180
Db	435	CCATGCAAGTGGTTAGCTTATTAAAGCAATGATGATCCCTTCAATTGCTCATCTTTC	494
OY	181	TGTTGGTGCAGCCCTTTGGCAGATGGAGCACTGGGTGTCATTCGATGGGGCACTCTTC	240
Db	495	TGTTGGTGCAGCCCTTTGGCAGATGGGGCACTGGGTGTCATTCGATGGGGCACTCTTC	554
OY	241	TGAAGATCTTGGGGCACTGTGTCACAGTGCATGACGTTGTTCACAGTGGGCACTTCC	300
Db	555	TGAAGATCTTGGGGCACTGTGTCACAGTGCATGACGTTGTTCACAGTGGGCACTTCC	614
OY	301	TCATCGCAGCGGGGCTTGCTTCTCTGCTTGGTTCCGCGGGCTGCTATGGGCTAAGA	360
Db	615	TCATCGCAGCGGGGCTTGCTTCTCTGCTTGGTTCCGCGGGCTGCTATGGGCTAAGA	674
OY	361	CTGAGAGCAAGTGTGCCCTGTGACGTTCTCTTCATCCCTCCTCATCTTCAATTGCTG	420
Db	675	CTGAGAGCAAGTGTGCCCTGTGACGTTCTCTTCATCCCTCCTCATCTTCAATTGCTG	734
OY	421	AGGTTGCAGTGTGCTGTGTGCTTGTGTGTACACCAATAGGCTGAGCACTTCTGAAGT	480
Db	735	AGGTTGCAGTGTGCTGTGTGCTTGTGTGTGTACACCAATAGGCTGAGCACTTCTGAAGT	794

Oy	481	TGCTGGTAGAGCCGCGCATCAAAAGATATATGCTTCCAGGAAGACTTCACTCAAGGT	540
Db	795	TGCTGGTAGAGTCCCTGCCATCAAAAGATATATGCTTCCAGGAAGACTTCACTCAAGGT	854
Oy	541	GAAGACACCACCATGTAAAGGGCTCAAGTGTGTGGCTTCCACCACTATACGGATTTTGGAG	600
Db	855	GAAGACACCACCATGTAAAGGGCTCAAGTGTGTGGCTTCCACCACTATATGGATTTTGGAG	914
Oy	601	ACTCAACCCCTACTTCAAGAGACAGTGTCTTCCCATCTTCTGTTCANATGACAAGTCA	660
Db	915	ACTCAACCCCTACTTCAAGAGACAGTGTCTTCCCATCTTCTGTTCANATGACAAGTCA	974
Oy	661	CCAAACAGCGCAATGAAAGCTGACCAAGCAAAAGGCTCAGACCAAAAGATAGAGGGTT	720
Db	975	CCAAACAGCGCAATGAAAGCTGACCAAGCAAAAGGCTCAGACCAAAAGATAGAGGGTT	1034
Oy	721	GCTTCAATCAAGCTTTTGTATGACATCCGAATATATCATATCACCGTGGGTGTGGAGAG	780
Db	1035	GCTTCAATCAAGCTTTTGTATGACATCCGAATATATCATATCACCGTGGGTGTGGAGAG	1094
Oy	781	CTGAAATTTGGGGGCTTCGAGCTGGTGCATATGTGTGCATATATCTGTACTGCAATC	840
Db	1095	CTGAAATTTGGGGGCTTCGAGCTGGTGCATATGTGTGCATATATCTGTACTGCAATC	1154
Oy	841	TACAAATAGATCACTTCTGCTCTGGCCTACTACGCTGCACATGGGAAGTGTGAAGAGGC	900
Db	1155	TACAAATAGATCACTTCTGCTCTGGCCTACTACGCTGCACATGGGAAGTGTGAAGAGGC	1214
Oy	901	ACCCTGGCAAGCAGACAGTATGTGGGGAGGGGACAGATCTTAACATATGACTTGGGCGCA	960
Db	1215	ACCCTGGCAAGCAGACAGTATGTGGGGAGGGGACAGATCTTAACATATGACTTGGGCGCA	1274
Oy	961	GAATGAGACCTGGCCCTTCTGCTCAGACTTGGGGCTAGATAGGAGCACTCCTTTTA-GC	1019
Db	1275	GAATGAGACCTGGCCCTTCTGCTCAGACTTGGGGCTAGATAGGAGCACTCCTTTTAAGGC	1334
Oy	1020	GATGCTGACTTCTTCTTCATATGTGGTGGTGGATGGGTGGGGGGCATTCAGAGCCTCTAA	1079
Db	1335	GATGCTGACTTCTTCTTCATATGTGGTGGTGGATGGGTGGGGGGCATTCAGAGCCTCTAA	1394
Oy	1080	GGTAGGCAAGTCTTGTGCCATTCGCCCAAGTCAATTAACCCCTGATATAGCCCTTAGGC	1139
Db	1395	GGTAGGCAAGTCTTGTGCCATTCGCCCAAGTCAATTAACCCCTGATATAGCCCTTAGGC	1454
Oy	1140	CTAGTGTGTATCCCAAGTGTCTACTAGGGGGATGAGAAAGGACATTTATATAGCTTGGGCA	1199
Db	1455	CTAGTGTGTATCCCAAGTGTCTACTAGGGGGATGAGAAAGGACATTTATATAGCTTGGGCA	1514
Oy	1200	TAAGTGAATATCAGACAGAGCCTGTGGGTGTATGTGTAAGAGCACTTCAAAATGCTATAAC	1259
Db	1515	TAAGTGAATATCAGACAGAGCCTGTGGGTGTATGTGTAAGAGCACTTCAAAATGCTATAAC	1574
Oy	1260	CTGTTACAAATGTTAAAAAATTTTTTTTTTTT 1289	
Db	1575	CTGTTACAAATGTTAAAAAATTTTTTTTTTTT 1604	
RESULT 5			
V34225 standard: DNA: 1324 BP.			
V34225; 28-JAN-1999 (first entry)			
DE Human secreted protein gene 4 clone HKCSR70.			
KW Human: secreted protein; fusion protein; gene therapy; protein therapy;			
KW diagnosis: tissue: cancer; tumour; neurodegenerative disorder; leukaemia;			
KW developmental abnormality; foetal deficiency; blood; allergy; renal; ds;			
KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;			
KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;			
KW congenital disorder; schizophrenia; prostate; obesity; osteoclast; thymus;			
KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;			
KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.			
OS Homo sapiens.			
PN W09839446-A2.			



D	290	TGAGATCTTGGGGCACA	TCGTCGTCAGTGGCAGATTTGCAACGTGGCTACTCC	349
O	301	TCATCGACGCGCGCTGTGGTCTTTGGCTCTTGGTTTC	CTGGGCTGCTAATGTTGAAGA	350
D	350	TCATCGCAGCGCGGCTGTGGTCTTTGGCTCTTGGTTTC	CTGGGCTGCTAATGTTGAAGA	409
O	361	CTGAGACCAATGTGGCTCTGTCAGCTTCTTCTTCAATCC	CTCTCTATCTTCAATTTGCTG	420
D	410	CTGAGAGCAATGTGGCTCTGTCAGCTTCTTCTTCAATCC	CTCTCTATCTTCAATTTGCTG	469
O	421	AGATTCACGCTGCTGGTCCCTTGGTACACACAAATGGCTGACACTTCC	TGACGT	480
D	470	AGGTTGCAGCTGCTGTGGTCCCTTGGTACACACAAATGGCTGACACTTCC	TGACGT	529
O	481	TGCTGTAGTGCCTGCCATCAAGAAAGTTATGGTTCC	ACGAGAACCTTCACTCAAGTGT	540
D	530	TGCTGTAGTGCCTGCCATCAAGAAAGTTATGGTTCC	ACGAGAACCTTCACTCAAGTGT	589
O	541	GGAACACCAACCATGAAAGGGCTCAAGTGTGTGGCTTCC	ACCACTATACGATTTTGAGG	600
D	590	GGAACACCAACCATGAAAGGGCTCAAGTGTGTGGCTTCC	ACCACTATACGATTTTGAGG	649
O	601	ACTGACCTTACTTGAAGAGAACAGTGGCTTCC	CCCATCTGTGTGCAATGACAAGTCA	660
D	650	ACTGACCTTACTTGAAGAGAACAGTGGCTTCC	CCCATCTGTGTGCAATGACAAGTCA	709
O	661	CCAACACAGCCCAATGAACCTCGACACAAAGGCTC	ACGACCAAAAGTAGAGGTT	720
D	710	CCAACACAGCCCAATGAACCTCGACACAAAGGCTC	ACGACCAAAAGTAGAGGTT	769
O	721	GCTTCATCAGCTTTTGTATGACATCCGAACCTAAT	GCATGACCGTGGTGTGTGGAG	780
D	770	GCTTCATCAGCTTTTGTATGACATCCGAACCTAAT	GCATGACCGTGGTGTGTGGAG	829
O	781	CTGGAATGGGGGGCTGAGCTGGCTGCGACAGATTTGT	GTCCATGATCTACTCTGCAATC	840
D	830	CTGGAATGGGGGGCTGAGCTGGCTGCGACAGATTTGT	GTCCATGATCTACTCTGCAATC	889
O	841	TACATAAAGTCCACTTTCGCTCTGCGCACTACTGCT	CCACATGGAACTGTGAAGAGC	900
D	890	TACATAAAGTCCACTTTCGCTCTGCGCACTACTGCT	CCACATGGAACTGTGAAGAGC	949
O	901	ACCCTGGCAGCAGCAGTGTGGGGGAGGGAGAGAGAT	CTAACCAATGTCACTTGGGCCA	960
D	950	ACCCTGGCAGCAGCAGTGTGGGGGAGGGAGAGAGAT	CTAACCAATGTCACTTGGGCCA	1000
O	961	GAATGGACCTCCCTTCTTGTGCTCCAGACTTGGGGCT	AGATNAGGACCACTTCCTTTAGCG	102
D	1010	GAATGGACCTCCCTTCTTGTGCTCCAGACTTGGGGCT	AGATNAGGACCACTTCCTTTAGCG	106
O	1021	ATGCTGACTTTCCTTCCTATGTGTGGGTGATGTGGGT	GGGGGCACTTCCAGAGCCTTTAAG	108
D	1070	ATGCTGACTTTCCTTCCTATGTGTGGGTGATGTGGGT	GGGGGCACTTCCAGAGCCTTTAAG	112
O	1081	GTAGCAGCTTCTGTTGCCATTTCCCCAGTCTATTTAA	ACCCTTGATATGCCCCCTAGGCC	114
D	1130	GTAGCAGCTTCTGTTGCCATTTCCCCAGTCTATTTAA	ACCCTTGATATGCCCCCTAGGCC	118
O	1141	TAGTGTGATATCCCAAGTGTCTACTTGGGGGATGAGA	AAAGGCAATTTATATAGCTGGGCA	120
D	1190	TAGTGTGATATCCCAAGTGTCTACTTGGGGGATGAGA	AAAGGCAATTTATATAGCTGGGCA	124
O	1201	AAGTGAATTCAGAGAGCCTCTGGGTGGATGTGTAGA	AGGCACTTCAAAATGCAATTAAC	126
D	1250	AAGTGAATTCAGAGAGCCTCTGGGTGGATGTGTAGA	AGGCACTTCAAAATGCAATTAAC	130
O	1261	TGTTACAAATGTTAA	1275	
D	1310	TGTTACAAATGTTAA	1324	
RESULT	6			
148113				

Query Match	Best Local Similarity	Matches 1284; Conservative	97.9%; 99.8%; 0;	Score 1262; Pred. No. 0;	DB 1; Mismatches 0;	Length 1288; Indels 2;	Gaps
OY	6	GGCCGTCCTGCGCCGCCACACAGCCGCGGAGCGTGTTCCTTTGTGGA	65				
DB	1	GCGCGTCCTGCGCCGCCACACAGCCGCGGAGCGTGTTCCTTTGTGGA	60				
OY	66	GCGTCAGCAGTTCCTCTTTCAGAACTCACTGCGCAAGAGCCCTGAAAGAGCACCACATG	125				
DB	61	GCGTCAGCAGTTCCTCTTTCAGAACTCACTGCGCAAGAGCCCTGAAAGAGCACCACATG	120				
OY	126	CAGTCCTTCACTTATTAAGACATGATGATCCTTTCATATTTGCTCATCTTTCTGTGT	185				
DB	121	CAGTCCTTCACTTATTAAGACATGATGATCCTTTCATATTTGCTCATCTTTCTGTGT	180				
OY	186	GGTGACGCGCTTGTGGCAGTGGGCACTGCGGTGCTAAATCGATGGGGCATCCTTTGTGAAG	245				
DB	181	GGTGACGCGCTTGTGGCAGTGGGCACTGCGGTGCTAAATCGATGGGGCATCCTTTGTGAAG	240				
OY	246	ATCTTCGGGCGCAGTCTGCGCAGTCCATGCGAGTTTGCAGAGTGGGCTACTTCCCTATC	305				
DB	241	ATCTTCGGGCGCAGTCTGCGCAGTCCATGCGAGTTTGCAGAGTGGGCTACTTCCCTATC	300				
OY	306	GCAGCCGGCGTGTGTGCTTGTGCTTGTGCTTGTGCTGAGGCTGCTATGCTAAGACTGAG	365				
DB	301	GCAGCCGGCGTGTGTGCTTGTGCTTGTGCTTGTGCTGAGGCTGCTATGCTAAGACTGAG	360				
OY	366	AGCAAGTGTGCGCTTGTGACAGTCTTCTTTCATCCCTCCTCATATTTCAATGCTAGAGTT	425				
DB	361	AGCAAGTGTGCGCTTGTGACAGTCTTCTTTCATCCCTCCTCATATTTCAATGCTAGAGTT	420				
OY	426	GCAGCTGTGTGCTGCGCTTGTGACAGCACAATGCGTGAAGCACTTCTGACGTTGCTG	485				
DB	421	GCAGCTGTGTGCTGCGCTTGTGACAGCACAATGCGTGAAGCACTTCTGACGTTGCTG	480				
OY	486	GTAAGTCTGCCATCAAGAAAGATTATGTTTCCAGAGAACTTCACTCAAGTGTGAGC	545				

Db 481 GTAGTGCCTGCCATCAAGAAAGATTATGTTCCAGAGAACTTCACTCAAGTGGAC 540  
 QY 546 ACCACCATGAAGGGGCTCAAGTGTGTGGCTTCCAAACTATACGGATTTTGGAGACTCA 605  
 Db 541 ACCACCATGAAGGGGCTCAAGTGTGTGGCTTCCAAACTATACGGATTTTGGAGACTCA 600  
 QY 606 CCTACTTCAAGAGAACAGTGCCTTCCCATCTCTGTTCATGATGACCAAGTCCACAC 665  
 Db 601 CCTACTTCAAGAGAACAGTGCCTTCCCATCTCTGTTCATGATGACCAAGTCCACAC 660  
 QY 666 ACAGCCATGAACCTTCAGCAGCAAAAGGCTCAGCAACAAAAGTAGAGGTTGCTTC 725  
 Db 661 ACAGCCATGAACCTTCAGCAGCAAAAGGCTCAGCAACAAAAGTAGAGGTTGCTTC 720  
 QY 726 AATCAGCTTTGTATGATCCGACAACTAATGACATCCAGTGGTGGTGGGACGCTGGA 785  
 Db 721 AATCAGCTTTGTATGATCCGACAACTAATGACATCCAGTGGTGGTGGGACGCTGGA 780  
 QY 786 ATTGGGGGCTCGAGTGGCTGGCATG-ATTGTGTCATGATCTGACTGACATCTACA 844  
 Db 781 ATTGGGGGCTCGAGTGGCTGGCATGATTTGTGTCATGATCTGACTGACATCTACA 840  
 QY 845 ATAAGTCCACTTGTGCTCTGCGCACTACTGCTGCACATGGGAATGTAAGAGCACCC 904  
 Db 841 ATAAGTCCACTTGTGCTCTGCGCACTACTGCTGCACATGGGAATGTAAGAGCACCC 900  
 QY 905 TGGCAACACAGATGATGGGGGGGAGGAGACAGATCTAACAATGTCATTTGGGCAAT 964  
 Db 901 TGGCAACACAGATGATGGGGGGGAGGAGACAGATCTAACAATGTCATTTGGGCAAT 960  
 QY 965 GGACCTGCTCTGTGCTGCTCCAGACTTGGGCTAGATAGGAGCACCTCTTTTA-GCGATG 1023  
 Db 961 GGACCTGCTCTGTGCTGCTCCAGACTTGGGCTAGATAGGAGCACCTCTTTTAGCGATG 1020  
 QY 1024 CTGACTTCTCTTCATTTGGTGGTGGATGGGAGGATTCAGAGCCTCTAAGGTA 1083  
 Db 1021 CTGACTTCTCTTCATTTGGTGGTGGATGGGAGGATTCAGAGCCTCTAAGGTA 1080  
 QY 1084 GCGAGTCTGTGGCCATTCGCCAGTCTATTAAAGCTTGATATGGCCCTTAGGCGCTAG 1143  
 Db 1081 GCGAGTCTGTGGCCATTCGCCAGTCTATTAAAGCTTGATATGGCCCTTAGGCGCTAG 1140  
 QY 1144 TGTGATCCAGTGTCTACTGTGGGATGAGAGAAAGCATTTTATAGCTGGGACATAAG 1203  
 Db 1141 TGTGATCCAGTGTCTACTGTGGGATGAGAGAAAGCATTTTATAGCTGGGACATAAG 1200  
 QY 1204 TGAATATGACAGACCTCTGGGTGGATGTGTAGAGGACCTTCAAAATGATAAACCTGT 1263  
 Db 1201 TGAATATGACAGACCTCTGGGTGGATGTGTAGAGGACCTTCAAAATGATAAACCTGT 1260  
 QY 1264 TACAATGTTAAAAAAGAAAAA 1289  
 Db 1261 TACAATGTTAAAAAAGAAAAA 1286

RESULT 7  
 ID V54014 standard: cDNA; 933 BP.  
 AC V54014;  
 DT 02-DEC-1998 (first entry)  
 DE Nucleotide sequence of the human tumour-associated antigen.  
 KM Human: tumour-associated antigen: PRAT: stimulation: cell proliferation;  
 KW antagonist: cancer: genetic defect: sickle cell anaemia: agonist;  
 OS antibody: hybridisation: probe: ss.  
 FT key location/Qualifiers  
 FT CDS 128..855  
 FT /tag- a  
 FT /product- "human tumour-associated antigen"  
 PD 03-SEP-1998.  
 PF 27-FEB-1998: U03953.  
 PR 28-FEB-1997: US-808148.

PA (INCY-) INCYTE PHARM INC.  
 PI Goli SK. Hillman JL;  
 DR WPI: 98-481208/41.  
 DR P-PSDB: W59954.  
 PT Human tumour-associated antigen PRAT - useful for stimulating cell  
 PT proliferation and screening for antagonists useful to treat or  
 PT prevent cell proliferation disorders e.g. cancers  
 PS Disclosure: Fig 1A-1C: 54pp: English.  
 CC This is the nucleotide sequence of the human tumour-associated antigen  
 CC (PRAT) used in the method of the invention for stimulating cell  
 CC proliferation and screening for antagonists useful to treat or prevent  
 CC cell proliferation disorders such as cancer, and genetic defect e.g.  
 CC sickle cell anaemia. The polypeptides can be combined with a suitable  
 CC carrier in pharmaceutical compositions, and also used to screen for  
 CC antagonists, agonists, and to generate antibodies. PRAT agonists can  
 CC be added to a cell to stimulate cell proliferation by increasing or  
 CC prolonging the activity of PRAT as above. The antagonists can be  
 CC combined with a suitable carrier in pharmaceutical compositions, which  
 CC can be administered to subjects to treat or prevent disorders  
 CC associated with cell proliferation, especially cancers. Antibodies  
 CC specific for PRAT may be used directly as antagonists, or indirectly  
 CC as a targeting or delivery mechanism to bring pharmaceutical agents to  
 CC PRAT-expressing cells. They are also useful to diagnose conditions or  
 CC diseases characterised by PRAT expression and to monitor therapeutic  
 CC interventions. The polynucleotide encoding PRAT, or complementary  
 CC sequences, can be used to produce hybridisation probes, useful to  
 CC detect polynucleotides or fragments encoding PRAT, e.g. to diagnose  
 CC diseases relating to polypeptide expression or monitor PRAT regulation  
 CC during therapeutic intervention.  
 SQ Sequence 933 BP: 198 A: 254 C: 236 G: 245 T:

Query Match 72.0%; Score 928; DB 1; Length 933;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-266;  
 Matches 928; Conservative 0; Mismatches 0; Indels 0; Gaps 0:

QY 1 AGCCAGGCGCCCTTGTGCTGCGCCACTAGTGAGCAACCCGGAGCCTTTTGTGCTTT 60  
 Db 6 AGCCAGGCGCCCTTGTGCTGCGCCACTAGTGAGCAACCCGGAGCCTTTTGTGCTTT 65  
 QY 61 GTGAGCCTCAGAGTTCCTCTTTCAGACTACTGCCAAGAGCCCTGAAGAGAGCA 120  
 Db 66 GTGAGCCTCAGAGTTCCTCTTTCAGACTACTGCCAAGAGCCCTGAAGAGAGCA 125  
 QY 121 CCATGAGTGTCTACCTTCATTAAGACATGATGATCTCTTTCATTTGCTCATCTTTC 180  
 Db 126 CCATGAGTGTCTACCTTCATTAAGACATGATGATCTCTTTCATTTGCTCATCTTTC 185  
 QY 181 TGTGTGTGAGAGCCCTGTGGAGTGGGATCGAGGTGTCATGATGGGATCTTTC 240  
 Db 186 TGTGTGTGAGAGCCCTGTGGAGTGGGATCGAGGTGTCATGATGGGATCTTTC 245  
 QY 241 TGAAGATCTTGGGCGACTGTGCTCCAGTGCAGTGTGTCAAGTGTGGCTACTTTC 300  
 Db 246 TGAAGATCTTGGGCGACTGTGCTCCAGTGCAGTGTGTGTCAAGTGTGGCTACTTTC 305  
 QY 301 TCATCCAGAGCCGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 360  
 Db 306 TCATCCAGAGCCGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTG 365  
 QY 361 CTAGAGCAAGTGTGCGCTGCTGAGCTTCTTCTCACTCTCCCTCACTCATTTGCTG 420  
 Db 366 CTAGAGCAAGTGTGCGCTGCTGAGCTTCTTCTCACTCTCCCTCACTCATTTGCTG 425  
 QY 421 AGTTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 Db 426 AGTTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 485  
 QY 481 TGTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 Db 486 TGTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545  
 QY 541 GGAACACCAACATGAAGGGCTCAAGTGTGCTGCTTCAACCACTATGAGATTTGAGG 600



```

Db      546 GGAACACACACACATGAAGGGCTCAGTGTGTGCTTACCAACTATACGGATTGAGG 605
Oy      601 ACTCACCCTACTCTCAAGAGACAGTGGCTTCCCATTTCTGTCAATGACAACTCA 660
Db      606 ACTCACCCTACTCTCAAGAGACAGTGGCTTCCCATTTCTGTCAATGACAACTCA 665
Oy      661 CCACACAGCCAAATGAACCTGCACCAAGCAAAAGGCTCAGCAACAAAAGTAGAGGTT 720
Db      666 CCACACAGCCAAATGAACCTGCACCAAGCAAAAGGCTCAGCAACAAAAGTAGAGGTT 725
Oy      721 GCTTCAATCAGCTTTTGTATGACATCCGAACATATGACAGTACCGTGGTGTGGCAG 780
Db      726 GCTTCAATCAGCTTTTGTATGACATCCGAACATATGACAGTACCGTGGTGTGGCAG 785
Oy      781 CTGGAATGGGGGCTCGACCTGGCTCCATGATTTGTCTCATGTATCTTACTGCAATC 840
Db      786 CTGGAATGGGGGCTCGACCTGGCTCCATGATTTGTCTCATGTATCTTACTGCAATC 845
Oy      841 TACAATAGCTCACTTCTGCTCTGCACACTGCTGCACATGGGAAGTGAAGAGGC 900
Db      846 TACAATAGCTCACTTCTGCTCTGCACACTGCTGCACATGGGAAGTGAAGAGGC 905
Oy      901 ACCCTGGCAAGCAGCAGTGTGGGGA 928
Db      906 ACCCTGGCAAGCAGCAGTGTGGGGA 933

```

```

RESULT 8
V58493/c
ID      V58493 standard; cDNA: 801 BP.
AC      V58493;
DT      08-DEC-1998 (first entry)
DE      3' fragment of prostate tumour specific gene J1-19.
KM      Prostate tumour specific gene; human; prostate cancer; detection;
KW      therapy; ss.
OS      Homo sapiens.
PN      WO9837418-A2.
PD      27-AUG-1998.
PF      25-FEB-1998: U03690.
PR      09-FEB-1998: US-904809.
PR      25-FEB-1997: US-806596.
PR      01-AUG-1997: US-904809.
PA      (CORI-) CORIXA CORP.
PI      Dillon DC, Xu J;
PI      MPI: 98-480805/41.
PT      Novel human prostate specific tumour protein and fragments - useful
PT      for detecting and treating prostate cancers
PS      Claim 1, Page 42; 141PP; English.
CC      This sequence represents a human prostate tumour specific gene, and can
CC      be used in the method of the invention. The method is for detecting
CC      prostate cancer comprises contacting a biological sample with an agent
CC      able to bind an immunogenic portion of a prostate protein (such as
CC      encoded by this sequence). An antibody which binds to an immunogenic
CC      portion of the prostate protein, and the method can be used to detect,
CC      monitor progression of, or treat prostate cancers. The antibody may
CC      also be conjugated to a therapeutic agent for use in therapy of prostate
CC      cancers.
SQ      Sequence 801 BP: 209 A: 208 C: 198 G: 164 T;

```

```

Query Match      46.7%: Score 602.4; DB 1: Length 801;
Best Local Similarity 91.7%: Pred. No. 1.5e-169;
Matches 714; Conservative 0; Mismatches 55; Indels 10; Gaps 8;
Oy      23 CCACCTAGTGGCAACACCCGGAGCTGTGTTGCTTGTGAGACCTCAGCAGTTCCTC 82
Db      779 CCATATCAGGGCAACACCCGGGGGTGTTTNTGNTTGGGGGGGCTTANACNNTTCCTT 720
Oy      83 TT-TGCAACTACTCTGCAAGAGCCCTGGAACAGAGCCACCATGCTAGCTTACGCTCA 141
Db      719 TTATCAGAAATTCATGCGCAAGAGCCCTGGAACAGAGGGCCACCATGCGTTCAG-TTCA 661

```

```

Oy      142 TTAAACATGATGATCCTTTCATTTCTCA-TCTTTCTGTGTGCTGAGCCCTGTG 200
Db      660 TTAAACATGATGATCCTTTCATTTCTCA-TCTTTCTGTGTGCTGAGCCCTGTG 601
Oy      201 GCAGTGGGATCTGGGTG-TCAATGATGGGCAATCC--TTTCTGAAGATCTTGGGCC- 257
Db      600 GCAGTGGGATCTGGGTG-TCAATGATGGGCAATCC--TTTCTGAAGATCTTGGGCC 541
Oy      257 ACTGTGCTCAGT-GGCATGACGTTTGTCAAGTGGGCTACTTCC--CATGGACCCGG 313
Db      540 ACTGTGCTCAGTGGGCAATGATTTGTCAAGTGGGCTACTTCC--CATGGACCCGG 481
Oy      314 CGTTGTGCTTGTGCTTGTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373
Db      480 CGTTGTGCTTGTGCTTGTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421
Oy      374 TGCCCTGTGAGCTTCTTCTTCACTCCTCCTCATCTTCTCATTTGCTGAGTTGCACTGC 433
Db      420 TGCCCTGTGAGCTTCTTCTTCACTCCTCCTCATCTTCTCATTTGCTGAGTTGCACTGC 361
Oy      434 TGTGTGCTGCTTGTGCTTACACCAATGGCTGAGCACTTCTGAGCTTCTGCTGCTGCTG 493
Db      360 TGTGTGCTGCTTGTGCTTACACCAATGGCTGAGCACTTCTGAGCTTCTGCTGCTGCTG 301
Oy      494 TGCCATCAAGAAAGATTTATGTTTCCAGAGAACTTCACTCAAGTGTGGAACACCACT 553
Db      300 TGCCATCAAGAAAGATTTATGTTTCCAGAGAACTTCACTCAAGTGTGGAACACCACT 241
Oy      554 GAAAGGGCTCAAGTGTGCTGCTTCCACCACTATACGATTTTGAAGACTCACCTACTT 613
Db      240 GAAAGGGCTCAAGTGTGCTGCTTCCACCACTATACGATTTTGAAGACTCACCTACTT 181
Oy      614 CAAGAGAACAGTGGCTTCCCATCTTGTGCTATGACATGACACGTACCAACAGCCAA 673
Db      180 CAAGAGAACAGTGGCTTCCCATCTTGTGCTATGACATGACACGTACCAACAGCCAA 121
Oy      674 TGAACCTGCACCAAGCAAAAGGCTCAGCAGCAAAAGATGAGAGGTGTGCTCAATCAGCT 733
Db      120 TGAACCTGCACCAAGCAAAAGGCTCAGCAGCAAAAGATGAGAGGTGTGCTCAATCAGCT 61
Oy      734 TTTGTATGACATCCGAACTATGACACTACCGTGGTGTGCTGAGCTGGAATTTGGG 792
Db      60 TTTGTATGACATCCGAACTATGACACTACCGTGGTGTGCTGAGCTGGAATTTGGG 2

```

```

RESULT 9
V61150/c
ID      V61150 standard; cDNA: 801 BP.
AC      V61150;
DT      06-JAN-1999 (first entry)
DE      3' cDNA sequence of prostate tumour clone J1-19.
KW      Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS      Homo sapiens.
PN      WO9837093-A2.
PD      27-AUG-1998.
PF      25-FEB-1998: U03492.
PR      09-FEB-1998: US-020956.
PR      25-FEB-1997: US-806099.
PR      01-AUG-1997: US-904804.
PA      (CORI-) CORIXA CORP.
PI      Dillon DC, Xu J;
PI      MPI: 98-609886/51.
PT      Polypeptides comprising immunogenic portions of prostate proteins -
PT      used in a vaccine for the treatment of prostate cancer
PS      Claim 3, Page 41; 130PP; English.
CC      The present sequence is a new DNA which encodes an immunogenic portion
CC      of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC      can be used as a vaccine for the treatment of prostate cancer. The DNA
CC      was identified by analysis of a subtracted cDNA library obtained by
CC      subtracting a prostate tumour cDNA expression library with a normal
CC      tissue cDNA library.
SQ      Sequence 801 BP: 209 A: 208 C: 198 G: 164 T;

```

Query Match 46.7% Score 602.4; DB 1; Length 801;  
Best Local Similarity 91.7%; Pred. No. 1.5e-169;  
Matches 714; Conservative 0; Mismatches 55; Indels 10; Gaps 8;

OY 23 CCACCTCAGTGGCAGACCCGGAGGCTGTTTTCCTTGTGAGGAGCCTCAGCAGTTCCTC 82  
DB 779 CCAATTCAGGGGACACCCGGGGGTGTTTNTCNTTTGGGGGGCTNANGNGTTCCTTT 720  
OY 83 TT-TGAGAACTCAGTCCCAAGAGCCCTGAAAGAGCCACCATGAGTCTTCACTTCA 141  
DB 719 TTTTGAATTCAGTCCCAAGAGCCCTGAAAGAGCCACCATGAGTCTTCACTTCA 661  
OY 142 TTAAGACCATGATGATCTCTTCAATTGCTCA-TCTTCTGTGTGAGCAGCCCTGTG 200  
DB 660 TTAAGACCATGATGATCTCTTCAATTGCTCAATTCTTCTGAGGAGCCCTGTG 601  
OY 201 GCATGGGAGCTGTGGTG-TCAATGATGGGAGATC--TTTCTGAAAGATCTTGGGCC- 257  
DB 600 GCAGTGGGAGCTGTGGTGATGATGATGGGAGATCCTTTCTGAAAGATNTTGGGCCA 541  
OY 257 ACTGTCGTCACAT-GCCATGCAAGTTTTCAGTGGGCTACTTCT--CATGGAGCCGG 313  
DB 540 ACTGTCGTCACATGCAAGTTTTCAGTGGGCTACTTCT--CATGGAGCCGG 481  
OY 314 GCTTGTGCTTGTGCTTGTGCTTGTGCTGCTGCTATGCTGCTAAGAGTGAAGAGAGT 373  
DB 480 CCTTGTGCTTGTGCTTGTGCTTGTGCTGCTGCTATGCTGCTAAGAGTGAAGAGAGT 421  
OY 374 TGCCCTCGTACGTTCTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 433  
DB 420 TGCCCTCGTACGTTCTTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 361  
OY 434 TGTGTGCTGCTGT 493  
DB 360 TGTGTGCTGCTGT 301  
OY 494 TGCCATCAAGAAAGATTTATGTTTCCAGAGAGACTTCACTCAAGTGTGGAACACCAT 553  
DB 300 TGCCATCAAGAAAGATTTATGTTTCCAGAGAGACTTCACTCAAGTGTGGAACACCAT 241  
OY 554 GAAAGGCTCAAGT 613  
DB 240 GAAAGGCTCAAGT 181  
OY 614 CAAAGAGAACTGCTTCCGCCATTCCTGTGCAATGCAAGTGTGCAAGTGTGCAAGTGT 673  
DB 180 CAAAGAGAACTGCTTCCGCCATTCCTGTGCAATGCAAGTGTGCAAGTGTGCAAGTGT 121  
OY 674 TGAACCTGACCAAGCAAAAGGCTCAGCAGCAAAAGAGAGGTTGCTTCAATCAGT 733  
DB 120 TGAACCTGACCAAGCAAAAGGCTCAGCAGCAAAAGAGAGGTTGCTTCAATCAGT 61  
OY 734 TTTGATGACATCCCACTAATGACGTACCGTGGGTGTGTGCAAGTGTGAATTTGGG 792  
DB 60 TTTGATGACATCCCACTAATGACGTACCGTGGGTGTGTGCAAGTGTGAATTTGGG 2

RESULT 10  
V58494  
ID V58494 standard; cDNA; 740 BP.  
AC V58494;  
DT 08-DEC-1998 (first entry)  
DE 5' fragment of prostate tumour specific gene J1-19.  
KM Prostate tumour specific gene; human; prostate cancer; detection;  
therapy; ss.  
OS Homo sapiens.  
PN M0837418-A2.  
PD 27-AUG-1998.  
PE 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI: 98-480805/41.  
PT Novel human prostate specific tumour protein and fragments - useful  
for detecting and treating prostate cancers  
PS Claim 1; Page 42; 14pp; English.  
CC This sequence represents a human prostate tumour specific gene, and can  
be used in the method of the invention. The method is for detecting  
prostate cancer comprises contacting a biological sample with an agent  
able to bind an immunogenic portion of a prostate protein (such as  
encoded by this sequence). An antibody which binds to an immunogenic  
portion of the prostate protein, and the method can be used to detect,  
monitor progression of, or treat prostate cancers. The antibody may  
also be conjugated to a therapeutic agent for use in therapy of prostate  
cancers.  
SQ Sequence 740 BP; 161 A; 200 C; 160 G; 194 T;

Query Match 44.3% Score 570.8; DB 1; Length 740;  
Best Local Similarity 89.7%; Pred. No. 3.6e-160;  
Matches 647; Conservative 0; Mismatches 63; Indels 11; Gaps 4;

OY 1 AGCCAGGCGTCCCTCTGCTGCTCCACTAGTGCAGACACCCGGAGCTGTTTCTCTT 60  
DB 6 AGCCAGGCGTCCCTCTGCTGCTCCACTAGTGCAGACACCCGGAGCTGTTTCTCTT 65  
OY 61 GTGAGGCTCAGAGTTCCTCTTCAAGTCACTCACTGCAAGAGCCCTGAAAGAGAGCA 120  
DB 66 GTGAGGCTCAGAGTTCCTCTTCAAGTCACTGCAAGAGCCCTGAAAGAGAGCA 125  
OY 121 CCATGAGTGTGAGTTCATTAAGACATGATGATCTCTTCAATTTGCTCATCTTTC 180  
DB 126 CCATGAGTGTGAGTTCATTAAGACATGATGATCTCTTCAATTTGCTCATCTTTC 185  
OY 181 TGTGTGTGACAGCCCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 240  
DB 186 TGTGTGTGACAGCCCTGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGT 245  
OY 241 TGAAGATTTGGGGGCACTGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 300  
DB 246 TGAAGATTTGGGGGCACTGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTG 305  
OY 301 TCATGCAAGCCGGGCTGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCT 360  
DB 306 TCATGCAAGCCGGGCTGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCTTGTGTGCT 365  
OY 361 CTGAGAGCAATGTGCTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420  
DB 366 CGAGAGCAAGTGTGCTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 425  
OY 421 AGTTGCAAGTGTGTGTGCTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480  
DB 426 AAGTTGCAAGTGTGTGTGCTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 485  
OY 481 TGTGTGTGTGCTGTGCATCAAGAAAGATTAT-GTTTCCAGAGAGTCTTCACTCAAGT 539  
DB 486 TGTGTGTGTGCTGTGCATCAAGAAAGATTATGTTTCCAGAGAGTCTTCACTCAAGT 545  
OY 540 TGAAGACCAACATGAAGAGG--CTCAAGTGTGTGCTTCAACCAATATACG--AT 593  
DB 546 TGAAGACCAACATGAAGAGGCTCAATTTCTGTGCTTCCCACTATATACGGAAT 605  
OY 594 TTTGAGGACTACCTCTTCAAGAGAGAGTGTG--CTTTTCCCACTCTTGTGCA 649  
DB 606 TTTGAAGANTNCCCTACTTCCAAAAAANANTGCTTTTCCCTTCTTGTGCA 665  
OY 650 TGACAGTCAACCAAGAGCAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 709  
DB 666 TGAAGAACTCCCAANAGCCCAATNAAAGTGTGCTTCCCTTCTTGTGCA 725  
OY 710 A 710  
DB 726 A 726

RESULT 11  
 ID V61151 standard; cDNA: 740 BP.  
 AC V61151;  
 DT 06-JAN-1999 (first entry)  
 DE 5' cDNA sequence of prostate tumour clone J1-19.  
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
 OS Homo sapiens.  
 PN M09837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998: U03492.  
 PR 09-FEB-1998: US-020956.  
 PR 25-FEB-1997: US-806099.  
 PR 01-AUG-1997: US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillion DC, Xu J;  
 DR WPI: 98-609886/51.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 used in a vaccine for the treatment of prostate cancer  
 PS Claim 3, Page 41-42; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 can be used as a vaccine for the treatment of prostate cancer. The DNA  
 CC was identified by analysis of a subtracted cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SQ Sequence 740 BP: 161 A; 200 C; 160 G; 194 T;

Query Match 44.3%; Score 570.8; DB 1; Length 740;  
 Best Local Similarity 89.7%; Pred. No. 3.6e-160;  
 Matches 647; Conservative 0; Mismatches 63; Indels 11; Gaps 4;

QY 1 AGCCAGGCGTCCCTCTGCTGCTCCACTAGTGGCAACCCGGAGCTGTTTGTCTTT 60  
 DB 6 AGCCAGGCGTCCCTCTGCTGCTCCACTAGTGGCAACCCGGAGCTGTTTGTCTTT 65  
 QY 61 GTGAGAGCTTCCAGTTCCTCTTTTCAAGACTCACTGCCAAGAGCCCTGAACAGAGCCA 120  
 DB 66 GTGAGAGCTTCCAGTTCCTCTTTTCAAGACTCACTGCCAAGAGCCCTGAACAGAGCCA 125  
 QY 121 CCATGAGTCTTCAAGTTCATTAAGACATGATGATCTCTTCATTTCTCATCTTTC 180  
 DB 126 CCATGAGTCTTCAAGTTCATTAAGACATGATGATCTCTTCATTTCTCATCTTTC 185  
 QY 181 TGTGTGTGAGCCCTGTTGGAGTGGGATCTGGGTGTCAATCGATGGGATCTTTC 240  
 DB 186 TGTGTGTGAGCCCTGTTGGAGTGGGATCTGGGTGTCAATCGATGGGATCTTTC 245  
 QY 241 TGAAGATCTTGGGGCCACTGTGCTCCAGTGCATGACAGTTTGTCAACGTGGGCTACTTC 300  
 DB 246 TGAAGATCTTGGGGCCACTGTGCTCCAGTGCATGACAGTTTGTCAACGTGGGCTACTTC 305  
 QY 301 TCATGCGAGCGGGGCTTGGTGTCTTTGCTTGTCCCTGGGGCTGTATGGTCTAAGA 360  
 DB 306 TCATGCGAGCGGGGCTTGGTGTCTTTGCTTGTCCCTGGGGCTGTATGGTCTAAGA 365  
 QY 361 CTGAGAGCAAGTGTGCCCTGTGACGTTCTTTCATCTCTCTCATCTTATTTCTG 420  
 DB 366 CTGAGAGCAAGTGTGCCCTGTGACGTTCTTTCATCTCTCTCATCTTATTTCTG 425  
 QY 421 AGGTTGAGCTGTGTGTGCTGCTTGGTGTACACACCAATGGCTGAGCACTTCTGACGT 480  
 DB 426 AAGTTGAGCTGTGTGTGCTGCTTGGTGTACACCAATGGCTGAGCACTTCTGACGT 485  
 QY 481 TGTGTGTGAGCTGTGTGCTGCTTGGTGTACACCAATGGCTGAGCACTTCTGACGT 539  
 DB 486 TGTGTGTGAGCTGTGTGCTGCTTGGTGTACACCAATGGCTGAGCACTTCTGACGT 545  
 QY 540 TGGAAACCAACCAATGAAGGG---CTCAAGTGTGTGCTTTCACCACTATACGG---AT 593  
 ||||||| ||||||| | | | ||||||| ||||||| |

DB 546 TGGAAACCAACCAATGAAGGGCTCCATTTCTGNTGCTTCCCACTATACCGAATT 605  
 QY 594 TTTGAGACTCAACCTTACTTCAAGAGACAGG----CCTTCCCACTTCTGTGCA 649  
 DB 606 TTTGAAAGANTCCCTTACTTCCAAAAAANNTGCTTTTCCCTTCTGTGCA 665  
 QY 650 TGCAACAGTCCACCAACAGCCCAATGAACCTGCACCAAGCAAAAGGCTCAGACCAAA 709  
 DB 666 TGAAGACNTCCCAANNCAGCAATTAATAAACTTCCCTCCNNCAAAAAGGNTCNCAACAAA 725  
 QY 710 A 710  
 DB 726 A 726

RESULT 12  
 ID V58490 standard; cDNA: 729 BP.  
 AC V58490;  
 DT 08-DEC-1998 (first entry)  
 DE 5' fragment of prostate tumour specific gene N1-1862.  
 KW Prostate tumour specific gene; human; prostate cancer; detection;  
 therapy; ss.  
 OS Homo sapiens.  
 PN M09837418-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998: U03690.  
 PR 09-FEB-1998: US-904809.  
 PR 25-FEB-1997: US-806596.  
 PR 01-AUG-1997: US-904809.  
 PA (CORI-) CORIXA CORP.  
 PI Dillion DC, Xu J;  
 DR WPI: 98-480805/41.  
 PT Novel human prostate specific tumour protein and fragments - useful  
 for detecting and treating prostate cancers  
 PS Claim 1, Page 40-41; 141pp; English.  
 CC This sequence represents a human prostate tumour specific gene, and can  
 be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprises contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC encoded by this sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 SQ Sequence 729 BP: 157 A; 203 C; 165 G; 191 T;

Query Match 43.5%; Score 560.8; DB 1; Length 729;  
 Best Local Similarity 94.9%; Pred. No. 3.3e-157;  
 Matches 650; Conservative 0; Mismatches 26; Indels 9; Gaps 7;

QY 1 AGCCAGGCGTCCCTCTGCTGCTCCACTAGTGGCAACCCGGAGCTGTTTGTCTTT 60  
 DB 2 AGCCAGGCGTCCCTCTGCTGCTCCACTAGTGGCAACCCGGAGCTGTTTGTCTTT 61  
 QY 61 GTGAGAGCTTCCAGTTCCTCTTTTCAAGACTCACTGCCAAGAGCCCTGAACAGAGCCA 120  
 DB 62 GTGAGAGCTTCCAGTTCCTCTTTTCAAGACTCACTGCCAAGAGCCCTGAACAGAGCCA 121  
 QY 121 CCATGAGTCTTCAAGTTCATTAAGACATGATGATCTCTTCATTTCTCATCTTTC 180  
 DB 122 CCATGAGTCTTCAAGTTCATTAAGACATGATGATCTCTTCATTTCTCATCTTTC 181  
 QY 181 TGTGTGTGAGCCCTGTTGGAGTGGGATCTGGGTGTCAATCGATGGGATCTTTC 240  
 DB 182 TGTGTGTGAGCCCTGTTGGAGTGGGATCTGGGTGTCAATCGATGGGATCTTTC 241  
 QY 241 TGAAGATCTTGGGGCCACTGTGCTCCAGTGCATGACAGTTTGTCAACGTGGGCTACTTC 300  
 DB 242 TGAAGATCTTGGGGCCACTGTGCTCCAGTGCATGACAGTTTGTCAACGTGGGCTACTTC 301  
 QY 301 TCATGCGAGCGGGGCTTGGTGTCTTTGCTTGTCTTCTGGGGCTGTATGGTCTAAGA 360  
 ||||||| ||||||| | | | ||||||| ||||||| |

```
Db 302 TCATGCAACCGCGGCTTGTGCTTAGCTCTAGGTTTCTTGCGCTGTATGCTGTAAAG 361
Oy 361 CTGAGAGCAAGTGTGCGCCCTGTAGAGTTCTTCTTCATCTCTCCATCTTCATTGCTG 420
Db 362 CTGAGAGCAAGTGTGCGCCCTGTAGAGTTCTTCTTCATCTCTCCATCTTCATTGCTG 421
Oy 421 AGGTTGACAGCTGTGTGCGCTTGTGTACACCAATGAGCTGAGCACTCTGAGCT 480
Db 422 AGGTTGACAGCTGTGTGCGCTTGTGTACACCAATGAGCTGAGCACTCTGAGCT 480
Oy 481 TGTCTGTATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 538
Db 481 TGTCTGTATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 540
Oy 539 G-TGGAAACACACATGAAAGGCTCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 596
Db 541 GTTGAACACACATGAAAGGCTCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 600
Oy 597 GAGGACTACCTTACTTCAAGAGACAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 653
Db 601 GAAGANTCA-CCTACTTCAAGAAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 659
Oy 654 AACGTCACCAACACAGCCATGAA 678
Db 660 AACGTCACCAACACAGCCATGAA 684
```

## RESULT 13

```
V61147
ID V61147 standard; cDNA: 729 BP.
AC V61147:
DT 06-JAN-1999 (first entry)
DE 5' cDNA sequence of prostate tumour clone N1-1862.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN WO9837093-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR MPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 40; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
of a prostate tumour protein. The encoded immunogen, or the DNA itself,
can be used as a vaccine for the treatment of prostate cancer. The DNA
was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
Sequence 729 BP; 157 A; 203 C; 165 G; 191 T;
```

```
Query Match 43.5%; Score 560.8; DB 1; Length 729;
Best Local Similarity 94.9%; Pred. No. 3.3e-157;
Matches 650; Conservative 0; Mismatches 26; Indels 9; Gaps 7;
```

```
Oy 181 TGTGTGTGACAGCCCTGTGTGACAGTGGCATCTGGTGTCTCATGTGATGGGCACTCTTC 240
Db 182 TGTGTGTGACAGCCCTGTGTGACAGTGGCATCTGGTGTCTCATGTGATGGGCACTCTTC 241
Oy 241 TGAAGATTTGGGGCACTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
Db 242 TGAAGATTTGGGGCACTGTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 301
Oy 301 TCATGCAACCGCGGCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 360
Db 302 TCATGCAACCGCGGCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 361
Oy 361 CTGAGAGCAAGTGTGCGCCCTGTAGAGTTCTTCTTCATCTCTCCATCTTCATTGCTG 420
Db 362 CTGAGAGCAAGTGTGCGCCCTGTAGAGTTCTTCTTCATCTCTCCATCTTCATTGCTG 421
Oy 421 AGGTTGACAGCTGTGTGCGCTTGTGTACACCAATGAGCTGAGCACTCTGAGCT 480
Db 422 AGGTTGACAGCTGTGTGCGCTTGTGTACACCAATGAGCTGAGCACTCTGAGCT 480
Oy 481 TGTCTGTATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 538
Db 481 TGTCTGTATGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 540
Oy 539 G-TGGAAACACACATGAAAGGCTCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 596
Db 541 GTTGAACACACATGAAAGGCTCAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 600
Oy 597 GAGGACTACCTTACTTCAAGAGACAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 653
Db 601 GAAGANTCA-CCTACTTCAAGAAAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT 659
Oy 654 AACGTCACCAACACAGCCATGAA 678
Db 660 AACGTCACCAACACAGCCATGAA 684
```

## RESULT 14

```
V58489/c
ID V58489 standard; cDNA: 751 BP.
AC V58489:
DT 08-DEC-1998 (first entry)
DE 3' fragment of prostate tumour specific gene N1-1862.
KW Prostate; tumour specific gene; human; prostate cancer; detection;
therapy; ss.
OS Homo sapiens.
PN WO9837418-A2.
PD 27-AUG-1998.
PR 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR MPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
for detecting and treating prostate cancers
PS Claim 1; Page 40; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
be used in the method of the invention. The method is for detecting
prostate cancer comprises contacting a biological sample with an agent
able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
portion of the prostate protein. An antibody which binds to an immunogenic
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
Sequence 751 BP; 188 A; 196 C; 167 G; 156 T;
```

```
Query Match 41.0%; Score 528.4; DB 1; Length 751;
Best Local Similarity 88.0%; Pred. No. 1.5e-147;
Matches 654; Conservative 0; Mismatches 75; Indels 14; Gaps 9;
```



---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 07:29:01 : Search time 165.01 Seconds  
(Without alignments)  
716.145 Million cell updates/sec

Title: US-09-030-606-111

Perfect score: 1289  
Sequence: 1 AGCCAGCGCTCCCTGCTGCT.....GTTAAAAA.....AAAAA 1289

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database :

Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/1na/PCRU59.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/1na/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.4	6.5	977	4	US-08-855-140-2
2	75.4	5.8	1151	4	US-08-807-044-2
3	62.8	4.9	1452	4	US-08-807-044-4
4	62.8	4.9	1452	5	PCT-US91-04986-1
5	56	4.3	687	1	US-08-254-493-2
6	56	4.3	1120	1	US-08-254-493-3
7	56	4.3	687	3	US-08-408-222B-2
8	56	4.3	1120	3	US-08-408-222B-3
9	50	3.9	7218	2	US-08-232-463-14
10	38.8	3.0	349	1	US-07-603-451A-4
11	38.8	3.0	349	1	US-08-060-822A-4
12	38.8	3.0	349	5	PCT-US94-05257-4
13	36.2	2.8	5975	1	US-08-404-354B-1
14	36.2	2.8	5975	2	US-08-314-083B-1
15	36.2	2.8	5975	2	US-08-435-675B-1
16	36.2	2.8	5975	2	US-08-336-257A-3
17	36	2.8	44377	4	US-08-804-227C-7
18	34	2.6	729	5	PCT-US94-01149-17
19	33.2	2.6	1943	3	US-08-631-200-11
20	33.2	2.6	1943	3	US-08-829-553-11
21	33.2	2.6	1943	4	US-08-927-267A-11
22	33.2	2.6	1943	4	US-08-936-707A-11
23	33.2	2.6	1943	4	US-08-936-707A-11
24	32.6	2.5	912	5	PCT-US94-01149-31
25	32.4	2.5	975	2	US-08-671-525B-9
26	32.4	2.5	975	2	US-08-672-109B-9
27	32.4	2.5	975	3	US-08-842-045-9
28	32.4	2.5	975	4	US-08-842-238-9
29	32.4	2.5	2574	4	US-08-677-794A-8
30	32	2.5	2098	5	US-08-153-848-41
31	32	2.5	2098	5	PCT-US93-11153-41
32	31.8	2.5	1037	3	US-08-463-195-1
33	31.8	2.5	1037	4	US-08-636-883-1
34	31.8	2.5	6318	4	US-08-808-793-1
35	31.8	2.5	501	4	US-08-483-695-28
36	31.8	2.5	501	4	US-07-965-285-38
37	31.8	2.5	1332	4	US-09-057-762-1

## ALIGNMENTS

38	31.2	2.4	2093	1	US-08-287-001A-1	Sequence 1, Appl1
39	31.2	2.4	1333	4	US-08-288-630-1	Sequence 1, Appl1
40	31.2	2.4	2093	5	PCT-US95-09941-1	Sequence 1, Appl1
41	31	2.4	3707	1	US-08-118-101A-1	Sequence 1, Appl1
42	31	2.4	2544	4	US-08-469-412A-6	Sequence 6, Appl1
43	30.8	2.4	1054	1	US-08-148-215A-1	Sequence 1, Appl1
44	30.6	2.4	2150	1	US-08-299-849B-24	Sequence 24, Appl1
45	30.6	2.4	2099	1	US-08-299-849B-25	Sequence 25, Appl1

RESULT 1  
US-08-855-140-2  
Sequence 2, Application US/08855140  
Patent No. 5834022  
GENERAL INFORMATION:  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Baudman, Olga  
APPLICANT: Goll, Surya K.  
APPLICANT: Guegler, Karl J.  
TITLE OF INVENTION: NEW TRANSMEMBRANE PROTEIN  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/855,140  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0296 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 977 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: MYOMOT01  
CLONE: 779308  
US-08-855-140-2

Query Match 6.58; Score 83.4; DB 4; Length 977;  
Best Local Similarity 52.08; Pred. No. 6.1e-16;  
Matches 248; Conservative 0; Mismatches 211; Indels 18; Gaps 2;

QY	121	CCATGACAGTGGTTCAGTTCATTAAGACATGATGATCCTTCAATTGCTCATCTTC	180
DB	126	CAGGAGGCTGCTTGGCTGCTTGAAGTACATGATGCTCTTCAATTGATATTCGCG	185
QY	181	TGTGTGTGACGCCCTGTGGCGACATGCGATCTGCTCATGATGCGGATCCTTTC	240

```

Db 186 TCATGCTGCTGGGCTGCTGGAGTGGGCACTGGCTCTCCGTGCCAAGCACTTTG 245
Oy 241 TGAAGATCTTCGGGGCCACTGCTGCGATGCGCATGAGTTGTCAAGTGGGCTACTCC 300
Db 246 CCACCTTC-----TCCCCAGCTTCCCTTGTTGTCTGCGAGCCAACTTG 230
Oy 301 TCATCGCAGCGGCGTGTGTGCTTTGCTTGTGTTCTGTGGCTGCTATGTGTAGA 360
Db 291 TCATCGCATAGGACCATTTGTATGTGTGAAGGGCTTCCTCGGCTGCTGGGGCCATCA 330
Oy 361 CTGAGACCAAGTGTGCTGCTGTGACGCTTCTTTCATCCCTCCCTCATTTCTGCTG 420
Db 351 AGAAGAAAGTGTGCTGCTGCTCCTCAGCTTTTTCATCGCTCTGTGCTACCTCTAGCAG 410
Oy 421 AGGTTGAGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 477
Db 411 AGCTGATCTTACCATCTCTTTCTTTCTTACATGACAGAGTGAAGCAAGCCACAGA 470
Oy 478 CGTTGCTGTAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 537
Db 471 AGGACCTGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 530
Oy 538 TGTGAACACCAACCATTAAGGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 594
Db 531 CCTGAACATCATCAGGCTGAGATGCGATGCTGTGTGCTGCTGCTGCTGCTGCTGCTG 587

```

RESULT 2  
US-08-807-044-2

```

; Sequence 2, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,044
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0224 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1151 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SCORNOT01
; CLONE: 663655
; US-08-807-044-2

```

```

Query Match 5.8%; Score 75.4; DB 4; Length 1151;
Best Local Similarity 55.9%; Pred. No. 1.8e-13;
Matches 189; Conservative 1; Mismatches 133; Indels 15; Gaps 2;

Oy 129 TGCATGAGTTTATTAAAGCATGATGATCCTTCATATTGCTCATCTTCTGTGTGT 188
Db 226 TGCCTCAGGCGGCTGAAGTACCTCATGCTTGCGCTTCACCTCTMTTCTGCGTGGAGGC 285
Oy 189 GCAGCCCTTGTGAGCATGGGCGCATCTGGGTGTCATGATGAGGCGATCCTTTCTGAATC 248
Db 286 TGTGGGTGCTGAGTGTGCGCATCTGGCT-----GGCGGCACAGAGGGAGGC 333
Oy 249 TTGGGGCCACTGCTGCGCCAGTGCATGCACTTTGTCAACGTGGGCTACTTCTCATTCGA 308
Db 334 TTGCGCAGCTGTCTC---TTCTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 390
Oy 309 GCGGCGTGTGTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 368
Db 391 ACCGGCGCTTGTGTATGCGCATGCGCTTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 450
Oy 369 AAGTGTCCCTGTGACGCTTCTTTCATCTCTCTCATCTTCTCATCTTCTCATCTGAGTTGCA 428
Db 451 AAGTGTCTCTGCTCCTCCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 510
Oy 429 GCTGCTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 466
Db 511 ATGCCATCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 548

```

RESULT 3

```

US-08-807-044-4
; Sequence 4, Application US/08807044
; Patent No. 5863735
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL HUMAN TRANSMEMBRANE 4 SUPERFAMILY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/807,044
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0224 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1452 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```



IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 180140  
US-08-807-044-4

Query Match 4.9% Score 62.8; DB 4; Length 1452;  
Best Local Similarity 51.2% Pred. No. 1.4e-09;  
Matches 173; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

257 ACTGTCGCCAGTGGCCATGAGTTGTCAAGTGGGCTACTTCCATCCAGCCGGCGT 316  
196 AGTCTCTTCATACCCCTCCCTCAGCTGGGCAATGTGTTGTCATCGTGGGCTC 255  
317 TGTGCTCTTCTCTGTTGCTGGGCTGATGCTAGAGTGGAGAGAGTGGC 376  
256 TATATATGATGATGCTTCTGCTTCTGCTGATGCTTCTGCTGATGCTGCT 315  
377 CCTGTCAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 436  
316 GCTTATGTCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 375  
437 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496  
376 CCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 432  
497 CATCAAGAAAGATGATGCTTCCAGGAGACTTCTACAGTGGAGACACCATGAA 556  
433 CAGCATCCACCGCTTACCACTCAGACAAATAGCACCAAGCAGCGTGGGCTCATCCAGTC 492  
557 AGGCTCAAGTGTGCTGCTTCCACCACTATACGAT 594  
493 ATTCTGACGATGTGTGTATTAATGCGACGATGATT 530

RESULT 4  
PCT-US91-04986-1  
Sequence 1, Application PC/TUS9104986

GENERAL INFORMATION:

APPLICANT: Seed, Brian

APPLICANT: Allen, Janet

APPLICANT: Aruffo, Alejandro

APPLICANT: Camerlin, David

APPLICANT: Laufer, Leander

APPLICANT: Oquendo, Carmen

APPLICANT: Simmons, David L.

APPLICANT: Stamenkovic, Ivan

APPLICANT: Stengelin, Siegfried

TITLE OF INVENTION: Rapid Immunoselection Cloning Method

NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:

ADDRESSEE: Greenlee & Associates

STREET: 5370 Manhattan Circle, Suite 201

CITY: Boulder

STATE: Colorado

COUNTRY: USA

ZIP: 80303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/04986

FILING DATE: 19910715

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/553,759

FILING DATE: 13-JUL-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/498,809

FILING DATE: 23-MAR-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/379,076  
FILING DATE: 13-JUL-1989  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/160,416

FILING DATE: 25-FEB-1988

ATTORNEY/AGENT INFORMATION:

NAME: Wall, Margaret M.

REGISTRATION NUMBER: 33,462

REFERENCE/DOCKET NUMBER: 11-88C

TELECOMMUNICATION INFORMATION:

TELEPHONE: 303-499-8080

TELEFAX: 303-499-8089

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1452 base pairs

TYPE: NUCLEIC ACID

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 74..733

PCT-US91-04986-1

Query Match 4.9% Score 62.8; DB 5; Length 1452;

Best Local Similarity 51.2% Pred. No. 1.4e-09;

Matches 173; Conservative 0; Mismatches 162; Indels 3; Gaps 1;

257 ACTGTCGCCAGTGGCCATGAGTTGTCAAGTGGGCTACTTCCATCCAGCCGGCGT 316  
196 AGTCTCTTCATACCTCCCTCAGCTGGGCAATGTGTTGTCATCGTGGGCTC 255  
317 TGTGCTCTTCTCTGTTGCTGGGCTGATGCTAGAGTGGAGAGAGTGGC 376  
256 TATATATGATGATGCTTCTGCTTCTGCTGATGCTTCTGCTGATGCTGCT 315  
377 CCTGTCAGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 436  
316 GCTTATGTCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 375  
437 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 496  
376 CCTGCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 432  
497 CATCAAGAAAGATGATGCTTCCAGGAGACTTCTACAGTGGAGACACCATGAA 556  
433 CAGCATCCACCGCTTACCACTCAGACAAATAGCACCAAGCAGCGTGGGCTCATCCAGTC 492  
557 AGGCTCAAGTGTGCTGCTTCCACCACTATACGAT 594  
493 ATTCTGACGATGTGTGTATTAATGCGACGATGATT 530

RESULT 5

US-08-254-493-2

Sequence 2, Application US/08254493

Patent No. 543986

GENERAL INFORMATION:

APPLICANT: IKEYAMA, SHUICHI

APPLICANT: KOYAMA, MASARU

APPLICANT: MIYAKE, MASAYUKI

APPLICANT: SENO, MASAHARU

TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND

NUMBER OF SEQUENCES: 7

PRODUCTION THEREOF

CORRESPONDENCE ADDRESS:

ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN

STREET: 130 WATER STREET

CITY: BOSTON

STATE: MASSACHUSETTS

COUNTRY: US

1 ZIP: 02109  
 2 COMPUTER READABLE FORM:  
 3 MEDIUM TYPE: Floppy disk  
 4 COMPUTER: IBM PC Compatible  
 5 OPERATING SYSTEM: PC-DOS/MS-DOS  
 6 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 7 CURRENT APPLICATION DATA:  
 8 APPLICATION NUMBER: US/08/254,493  
 9 FILING DATE:  
 10 CLASSIFICATION: 435  
 11 PRIOR APPLICATION DATA:  
 12 APPLICATION NUMBER: US 07/865552  
 13 FILING DATE: 09-APR-1992  
 14 PRIOR APPLICATION DATA:  
 15 APPLICATION NUMBER: JP 079996-1991  
 16 FILING DATE: 12-APR-1991  
 17 PRIOR APPLICATION DATA:  
 18 APPLICATION NUMBER: JP 085396-1991  
 19 FILING DATE: 17-APR-1991  
 20 PRIOR APPLICATION DATA:  
 21 APPLICATION NUMBER: JP 022321-1992  
 22 FILING DATE: 07-FEB-1992  
 23 ATTORNEY/AGENT INFORMATION:  
 24 NAME: RESNICK, DAVID S.  
 25 REGISTRATION NUMBER: 34235  
 26 REFERENCE/DOCKET NUMBER: 41777  
 27 TELECOMMUNICATION INFORMATION:  
 28 TELEPHONE: (617) 523-3400  
 29 TELEFAX: (617) 523-6440  
 30 TELEX: 200291 STRE UR  
 31 INFORMATION FOR SEQ ID NO: 2:  
 32 SEQUENCE CHARACTERISTICS:  
 33 LENGTH: 687 base pairs  
 34 TYPE: nucleic acid  
 35 STRANDEDNESS: double  
 36 TOPOLOGY: linear  
 37 MOLECULE TYPE: cDNA to genomic RNA  
 38 HYPOTHETICAL: NO  
 39 ANTI-SENSE: NO  
 40 ORIGINAL SOURCE:  
 41 ORGANISM: HUMAN  
 42 CELL TYPE: BREAST CARCINOMA  
 43 CELL LINE: ZR-75-1  
 44 US-08-254-493-2

	Query Match	4.3%	Score 56:	DB 1:	Length 687:	
	Best Local Similarity	49.0%:	Pred. No. 1.le-07:			
	Matches 149:	Conservative	0:	Mismatches 155:	Indels	0:
Gy	130 GCTTCAGCCTTCAATTAAGACCATGATGATCCCTTCAATTGGTCATCTTTGTGTGGTG	189				
Dd	17 GCACCAAGTCATCAAATACGTGCTGTGGATTAACTTCACTTCCTGGCTGCCGGGA	76				
Gy	130 CAGCCCTGTGGCAGTGGGCATCTGGGTGTCAAATCGATGGGGCATCCTTTCTGAAGACT	249				
Dd	77 TTGCTGCTCTTCCCATTTGGACATAATGGCTCCGATTCACACTCTCAGACCAGAGCATCTTGG	136				
Gy	250 TCGGGCCACTGTGCTGCACATGGCCATSCAATTTGTCAAGTGGGGCTACTTCATCATGGAG	309				
Dd	137 AGCAGAACAATAATATTAATTAATTCACGTTCTACACAGAGACTATTAATTTGATGGAG	196				
Gy	310 CCGGCGTTGTGGTCTTTCCTTGTGTTCTCGGGCGCTATGGTGTAAAGCTAGAGACA	369				
Dd	197 CCGGGGCCCTCATGATGCTGTGGGCTTCTGTGGCGTGTGGGGGCTGTGCAGAGATCCC	256				
Gy	370 AGTGTGCCCTGCTGACGTTCTTCTTCATCCTCTCTCATCTTCAATGCTGAGGTTGAG	429				
Dd	257 AGTGCATGCTGGAGACTGTTCTTCGGCTTCCTCTGTGTATATTCGCATTTGAATAGCTG	316				
Gy	430 CTMG	433				
Dd	317 CGCG	320				

```

RESULT      6
US-08-254-493-3
Sequence 3, Application US/08254493
Patent No. 543986
GENERAL INFORMATION:
APPLICANT: IKEYAMA, SHUICHI
APPLICANT: Koyama, MASARU
APPLICANT: Koyama, MASARU
APPLICANT: MIYAKE, MASAYUKI
APPLICANT: SEMOO, MASAHARU
TITLE OF INVENTION: MONOCLONAL ANTIBODY, POLYPEPTIDES AND
TITLE OF INVENTION: PRODUCTION THEREOF
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. CONLIN; DIKE, BRONSTEIN, ROBERTS & CUSHMAN
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,493
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/865552
FILING DATE: 09-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 0799966-1991
FILING DATE: 12-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 0853966-1991
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 022321-1992
FILING DATE: 07-FEB-1992
ATTORNEY/AGENT INFORMATION:
NAME: RESNICK, DAVID S.
REGISTRATION NUMBER: 34235
REFERENCE/DOCKET NUMBER: 41777
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 523-3400
TELEFAX: (617) 523-6440
TELEX: 200291 STRE UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1120 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
CELL TYPE: BREAST CARCINOMA
CELL LINE: ZR-75-1
FEATURE:
NAME/KEY: CDS
LOCATION: 112..795
FEATURE:
NAME/KEY: mat _peptide
LOCATION: 115..795
US-08-254-493-3

```

Query Match 4.38; Score 56; DB 1; Length 1120;



```
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/254,493
FILING DATE: 06-JUN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP-079996-1991
FILING DATE: 12-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP-085396-1991
FILING DATE: 14-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP-022321-1992
FILING DATE: 07-FEB-1994
ATTORNEY/AGENT INFORMATION:
NAME: Resnick, David S.
REGISTRATION NUMBER: 34,235
REFERENCE/DOCKET NUMBER: 41777-DIV
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1120 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to genomic RNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM: human
CELL TYPE: breast carcinoma
CELL LINE: ZR-75-1
FEATURE:
NAME/KEY: Coding Sequence
LOCATION: 112...795
OTHER INFORMATION: E Mat peptide
US-08-408-222B-3
```

```
Query Match 4.3%; Score 56; DB 3; Length 1120;
Best Local Similarity 49.0%; Pred. No. 1.4e-07;
Matches 149; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 130 GCTTACGCTTCAATTAAGACATGATGATCCCTTCATTTGCTCATCTTCTGTGGTG 189
DB 128 GCACCAAGTGCATCAAAATACCTGCTGCTGATTTAACTTCACTTGTGGCCGGGA 187
QY 190 CAGCCCTGTGGACATGCGCATCTGGGTGTCATGATGAGGCGCATCTTCTGAAGATCT 249
DB 188 TTGCTGCTCTTCCATGTGACATATGCTCCGATTCGACTCTCAGACCAAGACATCTTGG 247
QY 230 TCGGGCCACTGTCGTCACATGCGCATGCTTGTCAAGCTGGGCTACTTCTCATCGCAG 309
DB 248 AGCAAGAAACTAATATAATATATTCACGCTTCTACACAGAGATTAATATTCATCGAG 307
QY 310 CCGGGGTTGTGCTTCTTGTCTTGTTCCTGGGCTGCTCATGCTATGCTAGACTGAGAGA 369
DB 308 CCGGGGCTCTATGATGCTGTGGGCTTCTGCGCTGCGGGCTGTGCAAGAGATCC 367
QY 370 AGTGTGACCTCTGAGCTTCTTCTCATCTCTCTCATCTTCTCATGAGGTTGAG 429
DB 368 AGTGCATGCTGGAGCTGTCTTCTGCGCTTCTTGTGTATATTCGCAATTAAGATGCTG 427
QY 430 CTGC 433
DB 428 CCGC 431
```

```
RESULT 9
US-08-232-463-14
; Sequence 14, Application US/08232463
```

```
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgt-Fls
US-08-232-463-14
```

```
Query Match 3.9%; Score 50; DB 2; Length 7218;
Best Local Similarity 6.2%; Pred. No. 2.8e-05;
Matches 26; Conservative 217; Mismatches 177; Indels 0; Gaps 0;

QY 26 CTCAGTGCACACCCGGAGCTGTTGTCTTGTGAGCCCTCAGAGTTCCTCTT 85
DB 1038 CTGGCTCAGAGTGCAGGAGCTGCGATTTTTTTTTTTTTTTTTTTTTT 1097
QY 86 CAGACACTAGCCAGAGCCCTGAGAGCCACGATGCGATGCTCAGCTCATTA 145
DB 1098 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1157
QY 146 GACCATGATGATCTTCAATTTGCTCATCTTCTGTGTGAGGAGCCCTGTGGCAGT 205
DB 1158 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1217
QY 206 GGCATCTGGGTTGATACGATGGGCACTCTTCTGTAAGATCTTGGGCACTGCTC 265
DB 1218 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1277
QY 266 CAGTGCATGAGCTGTGCAAGGTGGCTACTCTCTCATGAGAGCGGCTGTGGCTT 325
DB 1278 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1337
QY 326 TGCTCTGTGTTCTGCGCTGATGCTAAGACTGAGAGCAAGTGTGCTCTGCTGAC 385
```



COMPUTER READABLE FORM:  
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 MB STORAGE  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: WORDPERFECT 5.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/05257  
FILING DATE: 11-MAY-1994  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: 08/060,822  
FILING DATE: May 11, 1993  
APPLICATION NUMBER: 07/991,200  
FILING DATE: December 15, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Rebecca Ralph Gaumond  
REGISTRATION NUMBER: 35,152  
REFERENCE/DOCKET NUMBER: CH-0408  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 349 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
PCT-US94-05257-4

Query Match 3.0%; Score 38.8; DB 5; Length 349;  
Best Local Similarity 52.5%; Pred. No. 0.012; Mismatches 77; Indels 0; Gaps 0;  
Matches 85; Conservative 0;

OY 1128 TGGCCCCCTAGGCTAGTGTGATCCAGTCTCTACTGGGGGATGAGAGAGCACTTTT 1187  
DB 184 TGGCACCCTGTGGGGGCGGACGTAATAATCTGTAGAAAGAGTAAGAGGCGCAAGAC 243  
OY 1188 ATAGCCTGGGCATAGGAAATGACGACGACCTGTGGGTGATGTGAGAGGCACTTCA 1247  
DB 244 ACAGCCGGGATCAATGCGGACGTAAGAACTGCGCCCTTTACAGAGGCTCTAAATTTAA 303  
OY 1248 AATGCAATAACCTGTACATGTATAAAAAAAAAAAAAA 1289  
DB 304 ACAGATAAATTTTGTCTCAAGTTAAAAA 345

RESULT 13  
US-08-404-354B-1  
Sequence 1, Application US/08404354B  
Patent No. 5618720  
GENERAL INFORMATION:  
APPLICANT: Ellis, Steven Bradley  
APPLICANT: Williams, Mark E.  
APPLICANT: Harpold, Michael Miller  
APPLICANT: Schwartz, Arnold  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/404,354B

FILING DATE: 15-FEB-1995  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/314,083  
FILING DATE: 28-SEP-1994  
APPLICATION NUMBER: US 07/914,231  
FILING DATE: 13-JUL-1992  
APPLICATION NUMBER: US 07/603,751  
FILING DATE: 08-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-53192  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-238-0999  
TELEFAX: 619-238-0062  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5975 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: Coding Sequence  
LOCATION: 79...5700  
OTHER INFORMATION:  
US-08-404-354B-1

Query Match 2.8%; Score 36.2; DB 1; Length 5975;  
Best Local Similarity 53.1%; Pred. No. 0.4; Mismatches 68; Indels 0; Gaps 0;  
Matches 77; Conservative 0;

OY 274 TGCAGTTGTCAACGCTGAGCTTCTTCATGCGACGCGGCTGTGCTTGTCTTG 333  
DB 344 TGGAGTACTTCTTCCACGCTTCTTCATGAGGCGCATGAGATCATCGGCTAG 403  
OY 334 GTTCTGGGCTGCTATGTGCTAGACTGAGAGCAAGTGTCCCTCGTAGCGTTCTTCT 393  
DB 404 GCTTCTGTTCCACGAGGCGCTTCTGCGAGCGGCTGGAAGCTGTGCTGATTCATCA 463  
OY 394 TCATCTCTCTCCATCTTCATTGC 418  
DB 464 TCGTCTTCTGCGGCTTTCACGCGC 488

RESULT 14  
US-08-314-083B-1  
Sequence 1, Application US/08314083B  
Patent No. 5686241  
GENERAL INFORMATION:  
APPLICANT: Ellis, Steven Bradley  
APPLICANT: Williams, Mark E.  
APPLICANT: Harpold, Michael Miller  
APPLICANT: Schwartz, Arnold  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: CALCIUM CHANNEL COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: CA  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:







GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 11:17:19 ; Search time 2493.51 Seconds  
(without alignments)  
1401.618 Million cell updates/sec

Title: US-09-030-606-111  
Perfect score: 1289  
Sequence: 1 AGCCAGCGCTGCTGCTT.....GTTAAAAA 1289

Scoring table: IDENTITY\_NUC

Searched: 3748650 seqs, 1355682445 residues

Database :

Pending Patents NA: \*  
1: /cgn2\_6/ptodata/2/pna/PCTUS9\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/pna/US07\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/pna/US080\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/pna/US081A\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/pna/US081B\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/pna/US081C\_COMB.seq:\*  
7: /cgn2\_6/ptodata/2/pna/US081D\_COMB.seq:\*  
8: /cgn2\_6/ptodata/2/pna/US082A\_COMB.seq:\*  
9: /cgn2\_6/ptodata/2/pna/US082B\_COMB.seq:\*  
10: /cgn2\_6/ptodata/2/pna/US082C\_COMB.seq:\*  
11: /cgn2\_6/ptodata/2/pna/US083A\_COMB.seq:\*  
12: /cgn2\_6/ptodata/2/pna/US083B\_COMB.seq:\*  
13: /cgn2\_6/ptodata/2/pna/US083C\_COMB.seq:\*  
14: /cgn2\_6/ptodata/2/pna/US084B\_COMB.seq:\*  
15: /cgn2\_6/ptodata/2/pna/US084C\_COMB.seq:\*  
16: /cgn2\_6/ptodata/2/pna/US085B\_COMB.seq:\*  
17: /cgn2\_6/ptodata/2/pna/US086\_COMB.seq:\*  
18: /cgn2\_6/ptodata/2/pna/US087A\_COMB.seq:\*  
19: /cgn2\_6/ptodata/2/pna/US087B\_COMB.seq:\*  
20: /cgn2\_6/ptodata/2/pna/US087C\_COMB.seq:\*  
21: /cgn2\_6/ptodata/2/pna/US088A\_COMB.seq:\*  
22: /cgn2\_6/ptodata/2/pna/US088B\_COMB.seq:\*  
23: /cgn2\_6/ptodata/2/pna/US088C\_COMB.seq:\*  
24: /cgn2\_6/ptodata/2/pna/US088D\_COMB.seq:\*  
25: /cgn2\_6/ptodata/2/pna/US089A\_COMB.seq:\*  
26: /cgn2\_6/ptodata/2/pna/US089B\_COMB.seq:\*  
27: /cgn2\_6/ptodata/2/pna/US089C\_COMB.seq:\*  
28: /cgn2\_6/ptodata/2/pna/US089D\_COMB.seq:\*  
29: /cgn2\_6/ptodata/2/pna/US089E\_COMB.seq:\*  
30: /cgn2\_6/ptodata/2/pna/US090A\_COMB.seq:\*  
31: /cgn2\_6/ptodata/2/pna/US090B\_COMB.seq:\*  
32: /cgn2\_6/ptodata/2/pna/US090C\_COMB.seq:\*  
33: /cgn2\_6/ptodata/2/pna/US091A\_COMB.seq:\*  
34: /cgn2\_6/ptodata/2/pna/US091B\_COMB.seq:\*  
35: /cgn2\_6/ptodata/2/pna/US091C\_COMB.seq:\*  
36: /cgn2\_6/ptodata/2/pna/US092B\_COMB.seq:\*  
37: /cgn2\_6/ptodata/2/pna/US092C\_COMB.seq:\*  
38: /cgn2\_6/ptodata/2/pna/US092D\_COMB.seq:\*  
39: /cgn2\_6/ptodata/2/pna/US092E\_COMB.seq:\*  
40: /cgn2\_6/ptodata/2/pna/US093A\_COMB.seq:\*  
41: /cgn2\_6/ptodata/2/pna/US093B\_COMB.seq:\*  
42: /cgn2\_6/ptodata/2/pna/US093C\_COMB.seq:\*  
43: /cgn2\_6/ptodata/2/pna/US093D\_COMB.seq:\*  
44: /cgn2\_6/ptodata/2/pna/US093E\_COMB.seq:\*  
45: /cgn2\_6/ptodata/2/pna/US093F\_COMB.seq:\*  
46: /cgn2\_6/ptodata/2/pna/US093G\_COMB.seq:\*  
47: /cgn2\_6/ptodata/2/pna/US093H\_COMB.seq:\*  
48: /cgn2\_6/ptodata/2/pna/US093I\_COMB.seq:\*  
49: /cgn2\_6/ptodata/2/pna/US093J\_COMB.seq:\*  
50: /cgn2\_6/ptodata/2/pna/US093K\_COMB.seq:\*  
51: /cgn2\_6/ptodata/2/pna/US093L\_COMB.seq:\*  
52: /cgn2\_6/ptodata/2/pna/US093M\_COMB.seq:\*  
53: /cgn2\_6/ptodata/2/pna/US093N\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1289	100.0	1289	US-09-020-747-111	Sequence 111, App
2	1289	100.0	1289	US-09-020-956-111	Sequence 111, App
3	1289	100.0	1289	US-09-030-606-111	Sequence 111, App
4	1289	100.0	1289	US-09-030-607-111	Sequence 111, App
5	1289	100.0	1289	US-09-288-946-111	Sequence 111, App
6	1277.6	99.1	1376	US-09-148-545-14	Sequence 14, App
7	1276.4	99.0	1437	US-09-359-922-5033	Sequence 5033, App
8	1274.8	98.9	1605	US-08-957-182-1	Sequence 1, App
9	1273.6	98.8	1324	US-09-148-545-82	Sequence 82, App
10	1262	97.9	1288	US-09-010-146-4	Sequence 4, App
11	1262	97.9	1288	US-09-010-146-50	Sequence 50, App
12	928	72.0	933	US-08-808-148-2	Sequence 2, App
13	613	47.6	631	US-08-740-274-2	Sequence 2, App
14	602.4	46.7	801	US-08-806-099-16	Sequence 16, App
15	602.4	46.7	801	US-08-806-596-16	Sequence 16, App
16	602.4	46.7	801	US-08-904-804-16	Sequence 16, App
17	602.4	46.7	801	US-08-904-809-16	Sequence 16, App
18	602.4	46.7	801	US-09-020-747-16	Sequence 16, App
19	602.4	46.7	801	US-09-020-956-16	Sequence 16, App
20	602.4	46.7	801	US-09-030-606-16	Sequence 16, App
21	602.4	46.7	801	US-09-030-607-16	Sequence 16, App
22	602.4	46.7	801	US-09-288-946-16	Sequence 16, App
23	570.8	44.3	740	US-08-806-099-17	Sequence 17, App
24	570.8	44.3	740	US-08-806-596-17	Sequence 17, App
25	570.8	44.3	740	US-08-904-804-17	Sequence 17, App
26	570.8	44.3	740	US-08-904-809-17	Sequence 17, App
27	570.8	44.3	740	US-09-020-747-17	Sequence 17, App
28	570.8	44.3	740	US-09-020-956-17	Sequence 17, App
29	570.8	44.3	740	US-09-030-606-17	Sequence 17, App
30	570.8	44.3	740	US-09-030-607-17	Sequence 17, App
31	570.8	44.3	740	US-09-288-946-17	Sequence 17, App
32	570.8	44.3	740	US-08-806-099-13	Sequence 13, App
33	560.8	43.5	729	US-08-806-596-13	Sequence 13, App
34	560.8	43.5	729	US-08-904-804-13	Sequence 13, App
35	560.8	43.5	729	US-08-904-809-13	Sequence 13, App
36	560.8	43.5	729	US-09-020-747-13	Sequence 13, App
37	560.8	43.5	729	US-09-020-956-13	Sequence 13, App
38	560.8	43.5	729	US-09-030-606-13	Sequence 13, App
39	560.8	43.5	729	US-09-030-607-13	Sequence 13, App
40	560.8	43.5	729	US-09-288-946-13	Sequence 13, App
41	554.4	43.0	631	US-06-069-957-74	Sequence 74, App
42	528.4	41.0	751	US-08-806-099-12	Sequence 12, App
43	528.4	41.0	751	US-08-806-596-12	Sequence 12, App
44	528.4	41.0	751	US-08-904-804-12	Sequence 12, App
45	528.4	41.0	751	US-08-904-809-12	Sequence 12, App



: NUMBER OF SEQUENCES: 178  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: SEED and BERRY LLP  
 : STREET: 6300 Columbia Center, 701 Fifth Avenue  
 : CITY: Seattle  
 : STATE: WA  
 : COUNTRY: USA  
 : ZIP: 98104  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent Release #1.0, Version #1.30  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/09/020,956  
 : FILING DATE: 09-FEB-1998  
 : CLASSIFICATION:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: MAKI, David J.  
 : REGISTRATION NUMBER: 31,392  
 : REFERENCE/DOCKET NUMBER: 210121.427C2  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: (206) 622-4900  
 : TELEFAX: (206) 682-6031  
 : INFORMATION FOR SEQ ID NO: 111:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 1289 base pairs  
 : TYPE: nucleic acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: cDNA  
 : ORIGINAL SOURCE:  
 : ORGANISM: Homo sapiens  
 : US-09-020-956-111

Query Match 100.0%; Score 1289; DB 30; Length 1289;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1289; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGCCAGGCGCTCTCTGCTGCCCTCCACATCAGTGGCAACCCGGGAGCTGTTTGTCTTT 60  
 DB 1 AGCCAGGCGCTCTCTGCTGCCCTCCACATCAGTGGCAACCCGGGAGCTGTTTGTCTTT 60  
 QY 61 GTGAGCGCTCAGCATGCCCTCTTCAAGTCACTGCGCAAGCCCTGGAACAGAGGCA 120  
 DB 61 GTGAGCGCTCAGCATGCCCTCTTCAAGTCACTGCGCAAGCCCTGGAACAGAGGCA 120  
 QY 121 CCATGCAAGTCTTCAAGTCAATTAAGACATGATGATCTCTTCAATTGCTCATCTTTC 180  
 DB 121 CCATGCAAGTCTTCAAGTCAATTAAGACATGATGATCTCTTCAATTGCTCATCTTTC 180  
 QY 181 TGTGTGTGTGAGCCCTGTGGCAGTGGCATCTGGGTGTCAATGATGGGCAATCTTTC 240  
 DB 181 TGTGTGTGTGAGCCCTGTGGCAGTGGCATCTGGGTGTCAATGATGGGCAATCTTTC 240  
 QY 241 TGAAGATCTTGGGGCAGATGCTGCAGTGCATGCAATGTCATGTAAGTGGGCACTTCC 300  
 DB 241 TGAAGATCTTGGGGCAGATGCTGCAGTGCATGCAATGTCATGTAAGTGGGCACTTCC 300  
 QY 301 TCATCGACAGCCGGGCTTGTGCTTCTTGTGCTTCTGGGCTGCTATGTTGCTAGA 360  
 DB 301 TCATCGACAGCCGGGCTTGTGCTTCTTGTGCTTCTGGGCTGCTATGTTGCTAGA 360  
 QY 361 CTGAGAGCAAGTGTGCCCTGTGAGTCTTCTTCATCCTCCTCTCATCTTCATCTGTG 420  
 DB 361 CTGAGAGCAAGTGTGCCCTGTGAGTCTTCTTCATCCTCCTCTCATCTTCATCTGTG 420  
 QY 421 AGTTGCAAGTGTGTGCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 DB 421 AGTTGCAAGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480  
 QY 481 TGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540

DB 481 TGTGTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 540  
 QY 541 GGAACACACCATTAAGAGAGGCTCAAGTGTGCTGCTTCCACCACTATACGATTTGAGG 600  
 DB 541 GGAACACACCATTAAGAGAGGCTCAAGTGTGCTGCTTCCACCACTATACGATTTGAGG 600  
 QY 601 ACTGACCTACTTCAAGAGAGAGGCTTCCCATCTGTTGTAATGACAAAGCTCA 660  
 DB 601 ACTGACCTACTTCAAGAGAGAGGCTTCCCATCTGTTGTAATGACAAAGCTCA 660  
 QY 661 CCAACAGAGCCCAATGAACCTGACCAAGCAAAAGGCTCAGCAACCAAAAGTAGAGGTT 720  
 DB 661 CCAACAGAGCCCAATGAACCTGACCAAGCAAAAGGCTCAGCAACCAAAAGTAGAGGTT 720  
 QY 721 GCTTCAATCAGCTTTGTATGACATCCGAACATAATGCAAGTACCGTGGTGTGCGAG 780  
 DB 721 GCTTCAATCAGCTTTGTATGACATCCGAACATAATGCAAGTACCGTGGTGTGCGAG 780  
 QY 781 CTGGAATTGGGGGCTGAGCTGCTGCCAATGATGTTGTCATGATCTGACTGCAATC 840  
 DB 781 CTGGAATTGGGGGCTGAGCTGCTGCCAATGATGTTGTCATGATCTGACTGCAATC 840  
 QY 841 TACAATAAGTCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 DB 841 TACAATAAGTCCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900  
 QY 901 ACCCTGGCAAGCAGCAGTATGTTGGGAGGGGAGACAGATCTAACATGTCTACTTGGGCA 960  
 DB 901 ACCCTGGCAAGCAGCAGTATGTTGGGAGGGGAGACAGATCTAACATGTCTACTTGGGCA 960  
 QY 961 GAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
 DB 961 GAATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020  
 QY 1021 ATGCTGACTTCTCTTCATGATGTTGGTGGATGATGATGATGATGATGATGATGATGATG 1080  
 DB 1021 ATGCTGACTTCTCTTCATGATGTTGGTGGATGATGATGATGATGATGATGATGATGATG 1080  
 QY 1081 GTAGCCAGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 DB 1081 GTAGCCAGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140  
 QY 1141 TAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 DB 1141 TAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1200  
 QY 1201 AAGTGAATCAGACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 DB 1201 AAGTGAATCAGACAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1260  
 QY 1261 TGTTCATGTTAAAAA 1289  
 DB 1261 TGTTCATGTTAAAAA 1289

RESULT 3  
 US-09-030-606-111  
 : Sequence 111, Application US/09030606  
 : GENERAL INFORMATION:  
 : APPLICANT: Xu, Jiangchun  
 : APPLICANT: Dillon, David C.  
 : TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHOD  
 : NUMBER OF SEQUENCES: 224  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: SEED and BERRY LLP  
 : STREET: 6300 Columbia Center, 701 Fifth Avenue  
 : CITY: Seattle  
 : STATE: WA  
 : COUNTRY: USA  
 : ZIP: 98104  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,606
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121, 428C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 1289 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-606-111

```

```

Query Match      100.0%: Score 1289: DB 30: Length 1289:
Best Local Similarity 100.0%: Pred. No.0:
Matches 1289: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

QY 1 AGCCAGGCGTCCCTCTGCTGCCACTGACAGTGGCAACCCGGGAGCTGTTTGTCTTT 60
DB 1 AGCCAGGCGTCCCTCTGCTGCCACTGAGTGGCAACCCGGGAGCTGTTTGTCTTT 60
QY 61 GTGAGAGCTTACAGTTCCTCTTTTGAAGTCACTGCTCCAGAGCCCTGAACAGAGCCA 120
DB 61 GTGAGAGCTTACAGTTCCTCTTTTGAAGTCACTGCTCCAGAGCCCTGAACAGAGCCA 120
QY 121 CCATGCGAGTCTTCAAGTCTATTAGACATGATGATCCCTTCAATTTGCTCATCTTTC 180
DB 121 CCATGCGAGTCTTCAAGTCTATTAGACATGATGATCCCTTCAATTTGCTCATCTTTC 180
QY 181 TGTGTGTGTGAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
DB 181 TGTGTGTGTGAGCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 240
QY 241 TGAAGATCTTCCGGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
DB 241 TGAAGATCTTCCGGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 300
QY 301 TCATGCGAGCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
DB 301 TCATGCGAGCGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 360
QY 361 CTGAGAGCAAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
DB 361 CTGAGAGCAAGTGTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 420
QY 421 AGGTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
DB 421 AGGTGTGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 481 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
DB 481 TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
QY 541 GGAACACACCATGAAAGGGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
DB 541 GGAACACACCATGAAAGGGCTCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 600
QY 601 ACTGACCTTACTTCAAGAGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660
DB 601 ACTGACCTTACTTCAAGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 660

```

```

QY 661 CCACACAGCCCAATGAAACCTGACCAAGCAAAAGGCTCAGACCAAAAGTAGAGGTT 720
DB 661 CCACACAGCCCAATGAAACCTGACCAAGCAAAAGGCTCAGACCAAAAGTAGAGGTT 720
QY 721 GCTTCATCAGCTTTTGTATGACATCCGAATATGACATCACCCTGGGTGTGTGCGAG 780
DB 721 GCTTCATCAGCTTTTGTATGACATCCGAATATGACATCACCCTGGGTGTGTGCGAG 780
QY 781 CTGGAATTTGGGGGCTCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 781 CTGGAATTTGGGGGCTCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
QY 841 TACAATTAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 TACAATTAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 ACCCTGCAAGCAGCAGTATTTGGGGAGGAGCAGATCTAACATGTCACTTGGGCCA 960
DB 901 ACCCTGCAAGCAGCAGTATTTGGGGAGGAGCAGATCTAACATGTCACTTGGGCCA 960
QY 961 GAATGACCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 GAATGACCTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
QY 1021 ATGCTGACTTTCCTTCCATTTGATGATGATGATGATGATGATGATGATGATGATG 1080
DB 1021 ATGCTGACTTTCCTTCCATTTGATGATGATGATGATGATGATGATGATGATGATG 1080
QY 1081 GTAGCCAGTTCCTGTTGCCATTCCTCCAGTCTATTAAACCTTGATATGCCCCCTAG 1140
DB 1081 GTAGCCAGTTCCTGTTGCCATTCCTCCAGTCTATTAAACCTTGATATGCCCCCTAG 1140
QY 1141 TAGTGTGATCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
DB 1141 TAGTGTGATCCCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
QY 1201 AAGTGAATCAGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1201 AAGTGAATCAGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
QY 1261 TGTTCACATGTTAAAAA 1289
DB 1261 TGTTCACATGTTAAAAA 1289

RESULT 4
US-09-030-607-111
; Sequence 111, Application US/09030607
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392

```





EARLIER APPLICATION NUMBER: 60/047,601  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,580  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,568  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,314  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,569  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,311  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,671  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,674  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,669  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,312  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,313  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,672  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,315  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/048,974  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/056,886  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,877  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,889  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,892

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047,595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/057,761  
EARLIER FILING DATE: 05-Sep-1997  
EARLIER APPLICATION NUMBER: 60/047,599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047,614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043,576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047,501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043,670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056,632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,654  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056,908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048,964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057,650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056,884  
EARLIER FILING DATE: 1997-08-22  
NUMBER OF SEQ ID NOS: 280  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 14  
LENGTH: 1376  
TYPE: DNA

Query Match 99.1%; Score 1277.6; DB 33; Length 1376;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1288; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

OY 1 AGCCAGGCGTCCCTGCGCTGCCCACACTGAGTGGCAACCCCGGAGCTGTTTGCTCTT 60  
|||  
Db 53 agccaggcgctccctgctgcccacactgagtggaaccccgagctgtttgtcctt 112  
|||  
OY 61 GTGAGCGCTCAGACAGTTCCTTTTTCAGAACTCAGTGGCCAGAGCCCTTGAACAGGACCA 120  
|||  
Db 113 gtgagcgctcagacagtctcctcttcagaaactcagtcgaagcgagccctgagacagagcca 172  
|||

OY	121	CGATGCAATGGCTTGAAGTCAATTAATGAACAGTAAAGTAAATCCCTTCATTTGGCATCTTTC	180
OY	122	CGATGCAATGGCTTGAAGTCAATTAATGAACAGTAAAGTAAATCCCTTCATTTGGCATCTTTC	180
DB	173	ccatgcagctgcttaagctctcttaagaacatgataccctctcaattctgcacatcttc	232
OY	181	TGTGTGTGTGACAGCCCTGTGTGGCAATGGGGCAATCTAGGGTGTCATATCATATGATGGGATCTTTC	240
DB	223	ctgtgtgtgtgtgacagccctctgttgacagtgggcaatctgggtgtaacatcgatgtagcatcttc	292
OY	241	TGAAATCTTCGGGCGCACATGTGCTGCAGTGCATGCAAGTATTGTCAACGTGGGCTACTTCC	300
DB	293	tgaagatctctcggcccaactcgcacagtcgcagtccagttctgcaacgtggtactctcc	352
OY	301	TTCATGCGACCGCGGGGTGTGTGTCTTTTGTCTTTTCCGTGGGCTGCTATGCTGCTAAGA	360
DB	353	tcattcgccagccgggtctgtgtgtctctgtctctgtctctgtctcctgtgtctatgtgtctaga	412
OY	361	CTGAGAGCAATGTGAGCCCTGCTGAGTCTTCTTCTTCATCCGCTCCCATCTTCATTGCG	420
DB	413	cttgagagaaagtgtgtccctctgtgacgtctctctctcaatctctctcaatctctcatctgt	472
OY	421	AGGTTCAGCTGTGTGTGTGCGCTTGGTGTACACCAATATGGCTGAGCACTTCTGAGGT	480
DB	473	aggttgcagctgt	532
OY	481	TGCTGTGTGTGTGCTTCGCCATCATAGAAAGATTATGTCTTCCAGAGAGCTTCACTCAAGT	540
DB	533	tgtctgt	592
OY	541	GGAAACACCAACCATATAAAGGGGTCAAGTGTGTGGCTTACCAACATATATAGAAATTTAGG	600
DB	593	ggaaacacacacatagaagaaggtcccaagtgctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	652
OY	601	ACTCACCTTACTTCAAAAGAGAAACAGTGCCTTTTCCCATCTTGTGTTCATATGACAAAGTCA	660
DB	653	actcaccttacttcaagaagagaagtgcttctcccatctgttgcaatgacaagctca	712
OY	661	CCAACACAGCCCAATGAAGAACTGCACCAAGCAAAAGGCTCACGACCAAAAAGTAGAGGTT	720
DB	713	ccaaacacagcccaatgaaaacctgcacccaagaacaaaggctcacgacccaacaaagtagaggtt	772
OY	721	GCTTCAATCACCTTTTGTATGACATCCGAACTATAGAGTACACCGGGGTGTGGGCG	780
DB	773	gcttcaaacacgctctgtatgaaacatccgaacataagcaagtcacagtggtgtgtgtgtgtgtgt	832
OY	781	CTGGAATTTGGGGGCGCTCGAGCTGCGTCCATATGATGTGTTCATATCTGTACTCTCAATC	840
DB	833	ctggaattgtgggtgcctcgtgagctgt	892
OY	841	TACATATAGTCACTTCTGCTCTGTCCACTACTGTCTGCACATGGAACTGTGAAGAGGC	900
DB	893	tacaataaagttcaactcttcgtcctctgcacatactgtgcacatggaactgtgaagagagc	952
OY	901	ACCCGGGCAACACACAGTATGTGGGGAGGGGACAGATCTTAACATGTCACTTGGGCA	960
DB	953	accctgggaagcagatgtatctgggggaagggaacgaatacacaatgtcacttgggcca	1012
OY	961	GAATGAGACTCCCTTTCTGTGCTCCAGACTTGGGGCTAGATAGGAGACACTCTCTTTTA-GC	1019
DB	1013	gaatgagactctccttctctctctcagagcttggggctcagataagggaacactctctttagc	1072
OY	1020	GATGCTGACTTTCCTTCATATGGTGGGTGATAGGTGGGGGCACTTCAGAGCCTCTAA	1079
DB	1073	gatgcctgaactctctctccatctgt	1133
OY	1080	GGTAGCCAGTTCTGTTCGCCATTTCCCCCACTTATTAACACCTTGATATAGCCCCCTAAGC	1139
DB	1133	ggtagccagttctgt	1192
OY	1140	CTAGTGTGTATCCAGAGTCTTACTTGGGGGATAGAGAAAGCATTTTATAGCTTGGGGA	1199
DB	1193	ctagt	1253
OY	1200	TAAGTGAATATAGCAGAGACCTCTTGGTGGATGTGTAGAAAGCACTTCAAAATGATATAAC	1259

```

Db      1253      taagtgaattcagaagagccctcgggctgtagtgtaaggacattcacaatgcataac 1312
OY      1260      CTGTTACATGTTAAAAAAAAAAAAAAAAAAAAA 1289
          |||||||:|||||
Db      1313      ctgttacatgctttraaaaaaaaaaaaaaaaa 1342

RESULT 7
US-09-359-922-5033/C
: Sequence 5033, Application US/09359922
: GENERAL INFORMATION:
: APPLICANT: Hyseq, Inc.
: TITLE OF INVENTION: NOVEL CONTIGS OBTAINED FROM VARIOUS CDNA
: TITLE OF INVENTION: LIBRARIES
: FILE REFERENCE: 20411-752CON1
: CURRENT APPLICATION NUMBER: US/09/359,922
: CURRENT FILING DATE: 1999-07-22
: EARLIER APPLICATION NUMBER: US 09/205,155
: EARLIER FILING DATE: 1998-12-03
: NUMBER OF SEQ ID NOS: 13203
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 5033
: LENGTH: 1437
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-359-922-5033

```

Query Match	99.0%;	Score 1276.4;	DB 68;	Length 1437;
Best Local Similarity	99.8%;	Pred. No. 0;		
Matches 1288;	Conservative	0;	Mismatches 1;	Indels 1;
			Gaps	1;

OY	1	AGCAGAGGCTCCCTCTCTCCCTGCCCCACTCAAGTGGCAACACCCGGGAGCTGTTTGTCCTT	60
Db	1298	AGCCAGGAGGCTCCCTCTCTCCCTGCCCCACTCAAGTGGCAACACCCGGGAGCTGTTTGTCCTT	1239
OY	61	GTGGAGCCTCAGCAGTATCCCTCTTTTCAAACTACTGCGCAAGAGCCCTGAACAGAGCA	120
Db	1238	GTGGAGCCTCAGCAGTATCCCTCTTTTCAAACTACTGCGCAAGAGCCCTGAACAGAGCA	1179
OY	121	CCATGCAGTGCTTCAAGCTTCATTAAAGCAATGATATCTCTTCATTATGCTCATCTTTC	180
Db	1178	CCATGCAGTGCTTCAAGCTTCATTAAAGCAATGATATCTCTTCATTATGCTCATCTTTC	1119
OY	181	TGTGTGTGTCAGACCTCTTTGGCAGTGGGCATCTGGGTGTCATTCATGGGGCATCTTTC	240
Db	1118	TGTGTGTGTCAGACCTCTTTGGCAGTGGGCATCTGGGTGTCATTCATGGGGCATCTTTC	1059
OY	241	TGAAGATCTTGGGGCCACTGTCGCCAGTGGCCATGTGAGTTTGCACAGTGGGGCTACTTC	300
Db	1058	TGAAGATCTTGGGGCCACTGTCGCCAGTGGCCATGTGAGTTTGCACAGTGGGGCTACTTC	999
OY	301	TGATCGAGCCGGGTGTGGTCTTTTGCTCTTGAGTTCTTGGGCTGCTATGGTGTGAAGA	360
Db	998	TGATCGAGCCGGGTGTGGTCTTTTGCTCTTGAGTTCTTGGGCTGCTATGGTGTGAAGA	939
OY	361	CTGAGAGCAAGTGTGCCCTGTGACGTTCTTTACTCTCTCTCATCTTCAATGCTG	420
Db	938	CTGAGAGCAAGTGTGCCCTGTGACGTTCTTTACTCTCTCTCTCATCTTCAATGCTG	879
OY	421	AGGTGACGCTGGTGGTGGCCCTGTGGTGAACCAACATGGCTGAGCACTTCGAGCT	480
Db	878	AGGTGACGCTGGTGGTGGCCCTGTGGTGAACCAACATGGCTGAGCACTTCGAGCT	819
OY	481	TGCTGTAGTGCCTGGCCATCAAAAGATTATGTTCCGAGAAAGCTTCACTCAAGTGT	540
Db	818	TGCTGTAGTGCCTGGCCATCAAAAGATTATGTTCCGAGAAAGCTTCACTCAAGTGT	759
OY	541	GGAAACCAACCATGAAGAGGCTCAAGTGCTGTGGCTTCCACCACTATAGGATTTTGAGG	600
Db	758	GGAAACCAACCATGAAGAGGCTCAAGTGCTGTGGCTTCCACCACTATAGGATTTTGAGG	699



```

Oy 601 ACTACCCCTACTTTCAAAAGAGAAAGTGCCTTCCCTCCCATCTGCTTCAGTGCAGACGTCA 660
Db 638 ACTCACCCCTACTTTCAAAAGAGAAAGTGCCTTCCCTCCCATCTGCTTCAGTGCAGACGTCA 639
Oy 661 CCAACACAGCCCAATGAAACCTGTCACACAAAGCAAAAGGCTCCACGACCAAAAAAGTAGAGGTT 720
Db 638 CCAACACAGCCCAATGAAACCTGTCACACAAAGCAAAAGGCTCCACGACCAAAAAAGTAGAGGTT 579
Oy 721 GCTTCATCAGCTTTTGTATGATCATTCCGAACATAATGCAGTACACCGTGGGTGTGTGCAG 780
Db 578 GCTTCATCAGCTTTTGTATGATCATTCCGAACATAATGCAGTACACCGTGGGTGTGTGCAG 519
Oy 781 CTGGAAATGGGGGGGCTGAGCTGGGCTGCCATGATGTGTCCATGATTCGTACTGCAATC 840
Db 518 CTGGAAATGGGGGGGCTGAGCTGGGCTGCCATGATGTGTCCATGATTCGTACTGCAATC 459
Oy 841 TACAATAAGTCCACTTTTGCTGCTCGCCACTACTGCTGCACATGGGAACTGTGAAGAGGC 900
Db 458 TACAATAAGTCCACTTTTGCTGCTCGCCACTACTGCTGCACATGGGAACTGTGAAGAGGC 399
Oy 901 ACCCTGCGCAAGCAGCAGTGTATGGGGAGGGGACAGAGATCTAACAATGTCTACTTTGGGCCA 960
Db 398 ACCCTGCGCAAGCAGCAGTGTATGGGGAGGGGACAGAGATCTAACAATGTCTACTTTGGGCCA 339
Oy 961 GAATGAGACCTCCCTCTTCTGCTCCAGACTTGGGGGCTAGATAGGAGACCATCTCTTTTA -GC 1011
Db 338 GAATGAGACCTCCCTCTTCTGCTCCAGACTTGGGGGCTAGATAGGAGACCATCTCTTTTAGGC 279
Oy 1020 GATGCGCTGACTTTCTCTTCATATGGTGGGTGGATGGGGTGGGGGGGCAATCCAGAGACCTCTTA 1072
Db 278 GATGCGCTGACTTTCTCTTCATATGGTGGGTGGATGGGGTGGGGGGGCAATCCAGAGACCTCTTA 219
Oy 1080 GGTAGCCAGTTCTGTGCCCCATTCGCCCACTATTAACCCCTTGATATGCCCCCTAGGC 1135
Db 218 GGTAGCCAGTTCTGTGCCCCATTCGCCCACTATTAACCCCTTGATATGCCCCCTAGGC 159
Oy 1140 CTATGGTGATGCCAGTGCTCTACTTGGGGGATGAGAGAAAGCATTTTATAGCTGTGGCA 1199
Db 158 CTATGGTGATGCCAGTGCTCTACTTGGGGGATGAGAGAAAGCATTTTATAGCTGTGGCA 99
Oy 1200 TAAGTGAATATAGCAGAGACCTCTGGTGTGATGTGTAGAAAGCATTCAAAATGATCAATAAC 1255
Db 98 TAAGTGAATATAGCAGAGACCTCTGGTGTGATGTGTAGAAAGCATTCAAAATGATCAATAAC 39
Oy 1260 CTGTTACATGTTAAAAAAGAAAAA 1289
Db 38 CTGTTACATGTTAAAAAAGAAAAA 9

RESULT 8
US-08-957-182-1
: Sequence 1, Application US/08957182
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallee, Edward R.
: APPLICANT: Racie, Lisa A.
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Spaulding, Vikki
: APPLICANT: Agostino, Michael
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
: NUMBER OF SEQUENCES: 39
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Genetics Institute, Inc.
: STREET: 87 Cambridgepark Drive
: CITY: Cambridge
: STATE: MA
: COUNTRY: U.S.A.
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/957,182
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1605 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-957-182-1

Query Match          98.9%  Score 1274.8;  DB 27;  Length 1605;
Best Local Similarity 99.8%;  Pred. No. 0;
Matches 1287;  Conservative 0;  Mismatches 2;  Indels 1;  Gaps 1;

1  AGCCAGGCGTCCCTTGCTGCTGCCCACTAGTGGCAACACCGGGAGCTGTTGTCTTT
Db  315  AGCCAGGCGTCCCTTGCTGCTGCCCACTAGTGGCAACACCGGGAGCTGTTGTCTTT
Qy   61  GTGAGCCTCAGCAGTCCCTCTTTTCAGAACTCACTGCCCAAGAGCCCTGAACAGAGCCA
Db  375  GTGAGAGCTTCAGACATCCCTCTTTCAGAACTCACTGCCCAAGAGCCCTGAACAGAGCCA
Qy  121  CCATCAGTGGCTTACGCTTTCATTAGACCATGATATCTCTCAATTGCTCATCTTTC
Db  435  CCATCAGTGGCTTACGCTTTCATTAGACCATGATATCTCTCAATTGCTCATCTTTC
Qy  181  TGTGTGTGTACACCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCTTTC
Db  495  TGTGTGTGTACACCCCTGTTGGCAGTGGGCATCTGGGTGTCAATCGATGGGCATCTTTC
Qy  241  TGAAGATCTTGGGGCAGTGTCTGTCAGTGCAGTGCAGTGTGTCAACGTGGGCTACTTCC
Db  555  TGAAGATCTTGGGGCAGTGTCTGTCAGTGCAGTGCAGTGTGTCAACGTGGGCTACTTCC
Qy  301  TCATGCGACCCGGGCTGTGTGCTCTTGTCTCTTGTGTTCTGTGGGCTGTATGCTCTAGA
Db  615  TCATGCGACCCGGGCTGTGTGCTCTTGTGCTCTTGTGTTCTGTGGGCTGTATGCTCTAGA
Qy  361  CTGAGAGCAAGTGTGACCCCTGAGAGTCTTGTATCTCTCTCCATCTCATTTGCGT
Db  675  CTGAGAGCAAGTGTGACCCCTGAGAGTCTTGTATCTCTCTCCATCTCATTTGCGT
Qy  421  AGGTTGCAAGCTGCTGTGTGCGCTTGTGTACACCAATGGCTGAGCACTTCTGACGT
Db  735  AGGTTGCAAGCTGCTGTGTGCGCTTGTGTACACCAATGGCTGAGCACTTCTGACGT
Qy  481  TGTGTGTGTGCTGTCCATCAAGAAAGATTATGTTCCAGGAGACTTCACTCAAGTGT
Db  795  TGTGTGTGTGCTGTCCATCAAGAAAGATTATGTTCCAGGAGAACTTCACTCAAGTGT
Qy  541  GGAACACCAACCTGAAGGGGCTCAAGTGTGAGTTCACCAACATATACGATTTTGAAG
Db  855  GGAACACCAACCTGAAGGGGCTCAAGTGTGAGTTCACCAACATATACGATTTTGAAG
Qy  601  ACTCAACCTACTTCAAGAGAACAGTGTGCTTCCCACTTCTGTTGCATATGACAAGTCA
Db  915  ACTCAACCTACTTCAAGAGAACAGTGTGCTTCCCACTTCTGTTGCATATGACAAGTCA
Qy  661  CCAACAGACCAATGAACCTTGCACCAAGCAAAAGGCTTACGACCAAAAAGTAGAGGTT

```

```
DB 975 CCAACACAGCCAAATGAACCTGCACCAAGCAAAAGCCTCAGCACAAAAAGTAGAGGTT 1034
QY 721 GCCTCATCGCTTTTGTATGACATCGAAGCTATGACAGCACCCTGGGGGTGGCCAG 780
DB 1035 GCCTTCATCGCTTTTGTATGACATCGAAGCTATGACAGCACCCTGGGGGTGGCCAG 1094
QY 781 CTGGAATTTGGGGGCTCGAGCTGCTGCCATGATTTGTCTCATGTATCTTACTGCAATC 840
DB 1095 CTGGAATTTGGGGGCTCGAGCTGCTGCCATGATTTGTCTCATGTATCTTACTGCAATC 1154
QY 841 TACAATTAAGTCCACTTCTGCTCTGCCACTACTGCTGCCACATGGGAAGCTGTGAAGAGC 900
DB 1155 TACAATTAAGTCCACTTCTGCTCTGCCACTACTGCTGCCACATGGGAAGCTGTGAAGAGC 1214
QY 901 ACCCTGGCAAGCAGCAGTGTATGGGGGAGGGGACAGGATTAACAAATGCTACTTGGGCCA 960
DB 1215 ACCCTGGCAAGCAGCAGTGTATGGGGGAGGGGACAGGATTAACAAATGCTACTTGGGCCA 1274
QY 961 GAATGACCTGCCCCCTTCTGCTCCAGACTTGGGGGCTAGATAGGACCACTCTTTTA-GC 1019
DB 1275 GAATGACCTGCCCCCTTCTGCTCCAGACTTGGGGGCTAGATAGGACCACTCTTTTAAGCC 1334
QY 1020 GATGCTGACTTTCCTTCCATTGTGGGTGATGGGTGGGGGCACTTCCAGACCTCTTAA 1079
DB 1335 GATGCTGACTTTCCTTCCATTGTGGGTGATGGGTGGGGGCACTTCCAGACCTCTTAA 1394
QY 1080 GGTAGCCAGTCTGTTGGCCCATTCGCCAGTCTATTAAACCTTGAATGATGCCCCCTAAGC 1139
DB 1395 GGTAGCCAGTCTGTTGGCCCATTCGCCAGTCTATTAAACCTTGAATGATGCCCCCTAAGC 1454
QY 1140 CTAGTGTGATCCAGTCTCTACTGCGGGGATGAGAAAGCATTTTATAGCCTGGGCA 1199
DB 1455 CTAGTGTGATCCAGTCTCTACTGCGGGGATGAGAAAGCATTTTATAGCCTGGGCA 1514
QY 1200 TAAGTGAATACACAGAGCTCTGGGTGATGTGTAGAAAGCACTTCAAAATGCATTAAC 1259
DB 1515 TAAGTGAATACACAGAGCTCTGGGTGATGTGTAGAAAGCACTTCAAAATGCATTAAC 1574
QY 1260 CTGTTCAATGTTAAAAAATTTTTTTTTT 1289
DB 1575 CTGTTCAATGTTAAAAAATTTTTTTTTT 1604

RESULT 9
US-09-148-545-82
; Sequence 82: Application US/09148545
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: P2001P1
; CURRENT APPLICATION NUMBER: US/09/148,545
; EARLIER FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600

; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
```

EARLIER APPLICATION NUMBER: 60/056, 893  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 630  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 878  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 662  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 872  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 882  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 637  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 903  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 888  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 879  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 880  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 894  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 911  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 636  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 874  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 910  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 864  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 631  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 845  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 892  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/047, 595  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/057, 761  
EARLIER FILING DATE: 05-Sep-1997  
EARLIER APPLICATION NUMBER: 60/047, 599  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 588  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 585  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 586  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 590  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 594  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 589  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 593  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/047, 614  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043, 578  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/043, 576  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/047, 501  
EARLIER FILING DATE: 1997-05-23  
EARLIER APPLICATION NUMBER: 60/043, 670  
EARLIER FILING DATE: 1997-04-11  
EARLIER APPLICATION NUMBER: 60/056, 632  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 664

EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 876  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 881  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 909  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 875  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 862  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 887  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/056, 908  
EARLIER FILING DATE: 1997-08-22  
EARLIER APPLICATION NUMBER: 60/048, 964  
EARLIER FILING DATE: 1997-06-06  
EARLIER APPLICATION NUMBER: 60/057, 650  
EARLIER FILING DATE: 1997-09-05  
EARLIER APPLICATION NUMBER: 60/056, 884  
EARLIER FILING DATE: 1997-08-22  
NUMBER OF SEQ ID NOS: 280  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 82  
LENGTH: 1324  
TYPE: DNA  
Query Match 98.8%; Score 1273.6; DB 33; Length 1324;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1273; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 AGCCAGGCGCTCCCTGCTGCTGCCACTCAGTGGCAACACCCGGAGCTGTTTGTCTTT 60  
Db 50 agccaggcgctccctgctgctgccactcagtggaacacccggagctgtttgtcttt 109  
QY 61 GTGAGACCTTCAGCAGTTCCTCTTTCAGAACTCAGTCCAGAGCCCTGAACAGAGCA 120  
Db 110 gtgagaccttcagcagttccctcttccagaactcagtcgaagccctgaacagagcca 169  
QY 121 CCATGAGTGTCTTACGCTTCATTAAAGACCATGATGATCTCTTCAATTGCTCATTTTC 180  
Db 170 ccatgagtgtcttcaagcttcaatgaacatgatatgctcttcaattgtctatcttc 229  
QY 181 TGTGTGTGTCAGCCCTGTTGGCATGTGGCATGTGGGTGTCATGATGGGCATCTCTTC 240  
Db 230 tgtgtgtgtgagccctgttggcagtggaatcctgggtgtcaatcgaatgggcatcttc 289  
QY 241 TGAAGATCTTCGGGCGGACGTGTGCTTTGCTTGGTTCTGCGCTCTATGTGTAGAGA 300  
Db 290 tgaagatcttcgggcccacgtgtgcttccagtcagtcagttgtcaacgttggctacttc 349  
QY 301 TCATCGACGCGGCGGCTGTGCTTTGCTTGGTTCTGCGCTCTATGTGTAGAGA 360  
Db 350 tcatcgacgcggcgctgtgtgtcttgcctctgttccctggctcgtatgtgtcaaga 409  
QY 361 CTAGAGCAAGTGTGCGCTTCGATGCTTCTTCACTCTCTCTCATCTTATGTCTG 420  
Db 410 ctgagagcaagtgtgctctgtgacgttcttcttccctccctccatcttcatgtctg 469  
QY 421 AGTTGACAGTGTGTCGTCGCTTGTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 480  
Db 470 agttgacagtcgt 529  
QY 481 TCGTGTAGTGTGCTGTCATCAGAAAGATTATGTTCCAGAGAGCTTCACTCAAGTGT 540  
Db 530 tgcgtgtagtcctgtccatcaagaagattatgttccagagagcttcaactcaagtgct 589  
QY 541 GGAACACCAACATGAAGAGGCTTAAGTGTGTGCTTACACACTATTAAGGATTTGAGG 600  
Db 590 ggaacacnccatgaagaaggcttaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 649  
QY 601 ACACACCTACTTCAAGAGAAAGAGTGTGCTTCCCATTCGTTGTCATGACACGTCA 660  
|||||

D	b	650	actaacctacttcaaaagagaacagagcgcttcccccattctggtgcaatgcaacgcyca	709
Q	y	661	ccaaacacagccatgaaacctgcacacgaacaaaggcttgcacacaaaatagtagggctt	720
D	b	710	ccaaacacagccaaatgaaacctgcacacgaacaaaggcttgcacacaaaatagtagggctt	769
Q	y	721	gcttcattcagcctttttatgacatccgaaactaatccatcacccgngggctgtggagag	780
D	b	770	gcttcaatcagccttttgcagacatccgaactaaagcagtcacgctgggtgtggagag	829
Q	y	781	ctggaaattggggggccttcgacactggctggccatgattgtgtcattgattgtactgcaatc	840
D	b	830	ctggaaattggggggccttcgacactggctggccatgattgtgtcattgattgtactgcaatc	889
Q	y	841	tacaaatagctcacttctgctcttgccactactgctgcacatggggaaactgtgaaagagc	900
D	b	890	tacaaatagctcacttctgctcttgccactactgctgcacatggggaaactgtgaaagagc	949
Q	y	901	accttggcgaacagacatgattgggggagggagacagagattctaacattgtcacttggcgca	960
D	b	950	accttggcgaacagcagtgatttgggggagggagacagagattctaacattgtcacttggcgca	1009
Q	y	961	gaatgagacatcgccctttctcctcagactgtggggcctgagatgagagacacactcttttagcg	1020
D	b	1010	gaatgagacatcgccctttctcctcagactgtggggcctgagatgagagacacactcttttagcg	1069
Q	y	1021	atgccttgaactttcccttcatttgggtgggtggatgggtggggggcattccagacactcttaag	1080
D	b	1070	atgccttgaactttcccttcatttgggtgggtggatgggtggggggcattccagacactcttaag	1129
Q	y	1081	gtaccacagttctgttggccatttcccccagctctttaaacccttggatgtgcccccttagagcc	1140
D	b	1130	gtaccacagttctgttggccatttcccccagctctttaaacccttggatgtgcccccttagagcc	1189
Q	y	1141	tagtgatgagatccacagctccttacttgggggagatgagaaagcattttatagcctggcgcat	1200
D	b	1190	tagtgatgagatccacagctccttacttgggggagatgagaaagcattttatagcctggcgcat	1249
Q	y	1201	aagtgaaatcagacagaccccttctgggtgagatgtgtagaaagcatttcaaaatgcataaac	1260
D	b	1250	aagtgaaatcagacagaccccttctgggtgagatgtgtagaaagcatttcaaaatgcataaac	1309
Q	y	1261	tgtttacaaatgtttaa 1275	
D	b	1310	tgttacaatgtttaa 1324	
RESULT 10				
US-09-010-146-4				
: Sequence 4, Application US/09010146				
: GENERAL INFORMATION:				
: APPLICANT: NI, et al.				
: TITLE OF INVENTION: Polynucleotides and Polypeptides Encoding				
: TITLE OF INVENTION: Receptors				
: NUMBER OF SEQUENCES: 63				
: CORRESPONDENCE ADDRESS:				
: ADDRESSEE: Human Genome Sciences, Inc.				
: STREET: 9410 Key West Ave				
: CITY: Rockville				
: STATE: MD				
: COUNTRY: US				
: ZIP: 20850				
: COMPUTER READABLE FORM:				
: MEDIUM TYPE: IBM PC compatible				
: OPERATING SYSTEM: PC-DOS/MS-DOS				
: SOFTWARE: Patent Release #1.0, Version #1.30				
: CURRENT APPLICATION DATA:				
: APPLICATION NUMBER: US/09/010.146				
: FILING DATE:				
: CLASSIFICATION:				
: PRIOR APPLICATION DATA:				
: APPLICATION NUMBER: WO US98/00959				

Query Match	Best Local Similarity	Matches 1284	Conservative	Mismatches	Indels	Gaps
6	97.9%	Score 1262;	DB 30;	Length 1288;		
1	99.8%	Pred. No. 0;				
66	97.9%	Score 1262;	DB 30;	Length 1288;		
61	99.8%	Pred. No. 0;				
126	97.9%	Score 1262;	DB 30;	Length 1288;		
121	99.8%	Pred. No. 0;				
186	97.9%	Score 1262;	DB 30;	Length 1288;		
181	99.8%	Pred. No. 0;				
246	97.9%	Score 1262;	DB 30;	Length 1288;		
241	99.8%	Pred. No. 0;				
306	97.9%	Score 1262;	DB 30;	Length 1288;		
301	99.8%	Pred. No. 0;				
366	97.9%	Score 1262;	DB 30;	Length 1288;		
361	99.8%	Pred. No. 0;				
426	97.9%	Score 1262;	DB 30;	Length 1288;		
421	99.8%	Pred. No. 0;				
486	97.9%	Score 1262;	DB 30;	Length 1288;		
481	99.8%	Pred. No. 0;				
546	97.9%	Score 1262;	DB 30;	Length 1288;		
541	99.8%	Pred. No. 0;				
606	97.9%	Score 1262;	DB 30;	Length 1288;		
601	99.8%	Pred. No. 0;				
666	97.9%	Score 1262;	DB 30;	Length 1288;		
661	99.8%	Pred. No. 0;				
726	97.9%	Score 1262;	DB 30;	Length 1288;		
721	99.8%	Pred. No. 0;				





## RESULT 13

US-08-740-274-2

: Sequence 2, Application US/08740274

: GENERAL INFORMATION:

: APPLICANT: Jacobs, Kenneth

: APPLICANT: McCoy, John

: APPLICANT: Lavallee, Edward

: APPLICANT: Racine, Lisa

: APPLICANT: Merberg, David

: APPLICANT: Treacy, Maurice

: APPLICANT: Spaulding, Vikki

: TITLE OF INVENTION: SECRETED PROTEINS AND POLYPEPTIDES

: NUMBER OF SEQUENCES: 49

: TITLE OF INVENTION: ENCODING THEM

: CORRESPONDENCE ADDRESS:

: ADDRESS: Genetics Institute, Inc.

: STREET: 87 Cambridgepark Drive

: CITY: Cambridge

: STATE: Massachusetts

: COUNTRY: U.S.A.

: ZIP: 02140

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/740,274

: FILING DATE:

: CLASSIFICATION:

: ATTORNEY/AGENT INFORMATION:

: NAME: Brown, Scott A.

: REGISTRATION NUMBER: 32,724

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (617) 498-8224

: TELEFAX: (617) 876-5851

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 631 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: double

: TOPOLOGY: linear

: MOLECULE TYPE: cDNA

US-08-740-274-2

Query Match 47.68; Score 613; DB 19; Length 631;

Best Local Similarity 98.94; Pred. No. 9e-175;

Matches 624; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

OY 156 ATCCCTTCATTTGTCATCTTCTGTCGTCAGCCCTGTCGAGTGGCATCTGG 215  
DB 1 ATCCCTTCATTTGTCATCTTCTGTCGTCAGCCCTGTCGAGTGGCATCTGG 60  
OY 216 GTGTCAATC-GATGGGATCTTCTGTAAGATCTTGGGCACTGTCTCAGTGGCAT 274  
DB 61 GTGTCAATCGAATGGGATCTTCTGTAAGATCTTGGGCACTGTCTCAGTGGCAT 120  
OY 275 GCAGTTGTCAAGCTGGCTACTTCTTCATCGACGCCGGGCTGTGCTTTGCTGTGG 334  
DB 121 GCAGTTGTCAAGCTGGGNNACTTCTTCATCGACGCCGGGCTGTGCTTTGCTGTGG 180  
OY 335 TTTCCTGGGCTGATGATGTAGTGAAGAGCAAGTGTGCTCTGAGTCTTCTT 394  
DB 181 TTTCCTGGGCTGATGATGTAGTGAAGAGCAAGTGTGCTCTGAGTCTTCTT 240  
OY 395 CATCTCTCTCTCATCTTTCATCTGCTGAGGTGCTGAGTGTGCTGCTGCTGTGCTAC 454  
DB 241 CATCTCTCTCTCATCTTTCATCTGCTGAGGTGCTGAGTGTGCTGCTGCTGTGCTAC 300  
OY 455 CACATAGGCTGAGCACTTCTGACGTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 514  
DB 301 CACATAGGCTGAGCACTTCTGACGTTGTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 360

## RESULT 14

US-08-806-099-16/c

: Sequence 16, Application US/08806099

: GENERAL INFORMATION:

: APPLICANT: Xu, Jiangchun

: TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS

: NUMBER OF SEQUENCES: 40

: CORRESPONDENCE ADDRESS:

: ADDRESS: Janet Sleath

: STREET: 1124 Columbia St, Ste464

: CITY: Seattle

: STATE: WA

: COUNTRY: USA

: ZIP: 98104

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patent Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/806,099

: FILING DATE: 24-FEB-1997

: CLASSIFICATION: 424

: ATTORNEY/AGENT INFORMATION:

: NAME: Makl, David J.

: REGISTRATION NUMBER: 31,392

: REFERENCE/DOCKET NUMBER: 210121.427

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 206-667-5728

: TELEFAX: 206-667-5728

: INFORMATION FOR SEQ ID NO: 16:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 801 base pairs

: TYPE: nucleic acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: MOLECULE TYPE: cDNA

US-08-806-099-16

Query Match 46.78; Score 602.4; DB 21; Length 801;

Best Local Similarity 91.78; Pred. No. 1.7e-171;

Matches 714; Conservative 0; Mismatches 55; Indels 10; Gaps 8;

OY 23 CCACTAGTGGCAACACCGGAGCTGTTTGTCTTGTGAGCCTCAGCAGTCCCTC 82  
DB 779 CCAATTAGGGCAACACCGGAGCTGTTTGTCTTGTGAGCCTCAGCAGTCCCTC 720  
OY 83 TT-TCAGACTCACTGCCAAGAGCCCTGAAACAGAGCCACCATGACGTGCTGACCTTA 141  
DB 719 TTTCAGAAATTCACCTGCCAAGAGCCCTGAAACAGAGCCACCATGACGTGCTGAC 651

QY 142 TTAACACCATGATGATCCCTTCATTTGGTCA-TCCTTGSTGTGGAGGCCCTGTG 200  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 660 TTAAACACCGANGATCCTTCTTCAATGGCTCATCTTTCNGNNGGCGACCCCTGTG 601  
  
QY- 201 GCAGTGGGCATCTGGGTG-TCAATCGATGGGGCATCC--TTTGAAGATCTTGGGCC- 257  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 600 GCAGTGGGCATNTGGGTGTTCATTCATCGATGGGGCATCCCTTTTCGAAAGATTGGGGCA 541  
  
QY 257 ACTGTCGTCACT-GCCATGCGAATTTGTCAACTGTGGGCTACTTCT--CATCGACCCGG 313  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 540 ACTGTCGTCACTGGCCATGCGAATTTGTCAACGTGGGTANTTTCTTCATTCGCGACGGG 481  
  
QY 314 CGTTGTGGCTTTGGCTTGGTTCTGCTGGGGCTGATAGTGTGAAGACTGAGACAAGTG 373  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 480 CGTTGTGGCTTTGGCTTGGTTCTGCTGGGGCTGATAGTGTGAAGACTGAGACAAGTG 421  
  
QY 374 TGCCCTGCTGACGTTCTTCTTCATCCCTCCTCATCTTCATTGCTGAGGTGGAGCTGC 433  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 420 TGCCCTGCTGACGTTCTTCTTCATCCCTCCTCATCTTCATTGCTGAGGTGGAGCTGC 361  
  
QY 434 TGTGCTGCGCTTGGTGTAAACCAATGGCTGAGCACTTCTGACGTTGCTGTGATGTC 493  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 360 TGTGCTGCGCTTGGTGTAAACCAATGGCTGAGCACTTCTGACGTTGCTGTGATGTC 301  
  
QY 494 TGCCATCAAGAAAGATTATGTTGCCAGAGAAGACTTCACTCAAGTGTGGAACACACCAT 553  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 300 TGCCATCAAGAAAGATTATGTTGCCAGAGAAGACTTCACTCAAGTGTGGAACACACCAT 241  
  
QY 554 GAAAGGCTCAAGTGTGCTGAGCTTCAACCAATATAGGATTTTGAGGACTCACCTACTT 613  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 240 GAAAGGCTCAAGTGTGCTGAGCTTCAACCAATATAGGATTTTGAGGACTCACCTACTT 181  
  
QY 614 CAAGAAGAACATGCTTCCCCCATCTGTGCAATGAGACAAGCTCACCAACAGCAA 673  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 180 CAAGAAGAACATGCTTCCCCCATCTGTGCAATGAGACAAGCTCACCAACAGCAA 121  
  
QY 674 TGAACCTGACCAAGCAAAAGGCTCAGACACAAAAAGTAGAGGTTGCTTCATAGCT 733  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 120 TGAACCTGACCAAGCAAAAGGCTCAGACACAAAAAGTAGAGGTTGCTTCATAGCT 61  
  
QY 734 TTTGATGACATCCGCACTAATGACATCACCGTGGGTGTGTGTGGACCTGGAAATGGGG 792  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 60 TTTGATGACATCCGCACTAATGACATCACCGTGGGTGTGTGTGGACCTGGAAATGGGG 2  
  
RESULT 15  
US-08-806-596-16/c  
Sequence 16, Application US/08806596  
GENERAL INFORMATION:  
APPLICANT: XU, Jiangchun  
TITLE OF INVENTION: COMPOUNDS FOR IMMUNODIAGNOSIS OF PROSTATE CANCER AND METHODS  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Janet Sleath  
STREET: 1124 Columbia St., Ste464  
CITY: Seattle  
STATE: WA  
COUNTRY: USA  
ZIP: 98104  
  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/806,596  
FILING DATE: 24-FEB-1997  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Makl, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.428

```

1 TELECOMMUNICATION INFORMATION:
2 TELEPHONE: 206-661-5728
3 TELEFAX: 206-667-5728
4 INFORMATION FOR SEQ ID NO: 16:
5 SOURCE CHARACTERISTICS:
6 LENGTH: 801 base pairs
7 TYPE: nucleic acid
8 STRANDEDNESS: single
9 TOPOLOGY: linear
10 MOLECULE TYPE: CDNA
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
10
```

Query Match	46.7%;	Score 602.4;	DB 21;	Length 801;
Best Local Similarity	91.7%;	Pred. No. 1.7e-171;		
Matches 714;	Conservative	0;	Mismatches 55;	Indels 10; Gaps 8;

QY	23	CAAGTCAGTGGCAACACCCGGGAGCGTTTGTGCTTTGTGGAGCCTCAGACAGTTCCCTC	82
Dp	779	CCATTCAGGGCAACACCCGGGGGGTGTTTTCTCTTTGGGGGGGGCTNANGCAGTTCTTT	722
QY	83	TT-TCAGAACTCACTGCCAAGAGCCCTGTAACAGAGGCCACCAATGACAGTCTTCACTTCA	14
Dp	719	TTTNTCAGAAATTCACGTGCCAAGAGCCCTGTAACAGAGGGGCCACCAATGCAAGTTTCAG-TCGA	66
QY	142	TTTAAGACCATGATGATCCTCTTCAATTGGTCTCA-TCCTTTGTGTGTGTACACCTGTTC	20
Dp	660	TTAAGACACANGANGATCCCTTCTCAATGGCTCAATCTTTGNGNGNGNAGCAGCCCTGTNG	60
QY	201	GCAGTGGGCACTCGGGTG-TCATGATGAGGGGCAATCC--TTTGTGAAGATCTTGGGGC-	25
Dp	600	GCAGTGGGCAATNGGGGTGTCAATGATGAGGGCAATCCCTTTTCTGAGATNTTGGGGCA	54
QY	257	ACTGTGTCACAGT-GGCATGCAAGTTGTTCACAGTGGGCTACTTCCP--CATGCGACGGG	31
Dp	540	ACGTGTGTCACAGTGGGCATGACAGTGTGTTCACAGTGGGTATTTCTTCAATGGCAGCNG	48
QY	314	CGTTGTGTCTTGGCTCTTGGTTTCGTGGCTGCTATGTGCTTAAGACTGAAGCAAGT	37
Dp	480	CGTTGTGTCTTGGCTCTTGGTTTCGTGGCTGCTATGTGCTTAAGACTGAAGCAAGT	42
QY	374	TGGCCCTGAGAGCTTCTTCATCCCTCCCTCATCTTCATTTAGTGGAGTTGGACGTCG	43
Dp	420	TGCCCTGAGAGCTTCTTCATCCCTCCCTCATCTTCATTTAGTGGAGTTGGACGTCG	36
QY	434	TGTGTGCGCTTGTGTACACACATGGCTGAGACATTTCTGACGTTCTGTGTAGTCC	49
Dp	360	TGTGTGCGCTTGTGTGTACACACATGGCTGAGACATTTCTGACGTTCTGTGTAGTCC	30
QY	494	TGCCATCAGAAAGATTATGTGTCCAGGAAGACTTCAATGATGTGAGACACCAAT	55
Dp	300	TGCCATCAGAAAGATTATGTGTCCAGGAAGACTTCAATGATGTGAGACACCAAT	24
QY	554	GAAAGGGCTCAAGTGTGGGTTCCACCAATATAGGATTTTGAAGATCCACCTACT	61
Dp	240	GAAAGGGCTCAAGTGTGGGTTCCACCAATATAGGATTTTGAAGATCCACCTACT	18
QY	614	CAAGAGAACAGTGTCTTCCCTCATTTCTGTGCAATGATCAACGTCACCAACAGCCAA	67
Dp	180	CAAGAGAACAGTGTCTTCCCTCATTTCTGTGCAATGATCAACGTCACCAACAGCCAA	12
QY	674	TGAACCTGCACCAAGCAAAAGGCTACAGACCAAAAAGTGAAGGTTGCTTAATACGT	73
Dp	120	TGAACCTGCACCAAGCAAAAGGCTACAGACCAAAAAGTGAAGGTTGCTTAATACGT	61
QY	734	TTTGTATGACATCCGAACATATGACATACCGTGGGTGTGTGGACGTGGAAATGGGG	792
Dp	60	TTTGTATGACATCCGAACATATGACATACCGTGGGTGTGTGGACGTGGAAATGGGG	2

Search completed: September 25, 1999, 11:17:51  
Job time: 8526 sec



---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:04:51 ; Search time 1811.29 Seconds  
(without alignments)  
1403.749 Million cell updates/sec

Title: US-09-030-606-111

Perfect score: 1289  
Sequence: 1 ACCCAGCGCTCCTCTGCCT.....GTTAAAAAAAAAAAAAAAAA 1289

Scoring table: IDENTITY\_NUC  
Searched: 2546578 seqs, 98626752 residues

Database : EST.\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	621.6	48.2	697	49	A1660579	A1660579 we68d03.x
2	545	42.3	593	28	AA100799	AA100799 zm26d01.s
3	518	40.2	557	51	A1708181	A1708181 as89f08.x
4	493.6	38.3	562	51	A1720370	A1720370 as79g12.x
5	478.2	37.1	607	36	AA659332	AA659332 nui3d01.s
6	477.4	37.0	479	33	AA420789	AA420789 nc63b06.x
7	464.8	36.1	505	34	AA468417	AA468417 nc78e09.r
8	458	35.5	470	35	AA593864	AA593864 n19f08.s
9	456.8	35.4	488	27	AA037886	AA037886 z104h10.s
10	444.4	34.5	446	39	AA846269	AA846269 a183h06.s
11	438	34.0	450	42	A1075023	A1075023 og97h07.x
12	437.6	33.9	453	36	AA659579	AA659579 nt63b06.s
13	434.2	33.7	496	34	AA534171	AA534171 n170b07.s
14	431.4	33.5	456	39	AA861869	AA861869 BK44806.s
15	428	33.2	440	48	A1568832	A1568832 th16g06.x
16	426.4	33.1	454	29	AA132967	AA132967 zo22b06.s
17	425	33.0	425	35	AA552617	AA552617 nk12c03.s
18	421	32.7	433	39	AA838513	AA838513 ce90g12.s
19	421	32.7	442	44	A1276101	A1276101 q174h03.x
20	420.4	32.6	434	45	A1393647	A1393647 t959f12.x
21	420	32.6	432	29	AA161043	AA161043 zo74g11.s
22	418	32.4	430	41	A1041400	A1041400 ow35c04.s
23	417	32.4	497	28	AA100852	AA100852 zm26d01.r
24	411.4	31.9	448	38	AA812351	AA812351 nr84b08.s
25	409.6	31.8	437	35	AA558636	AA558636 n159d12.s
26	409.4	31.8	412	29	AA132677	AA132677 zo20g06.r
27	402.4	31.2	426	29	AA146606	AA146606 zo35c09.s
28	400	31.0	403	35	AA574382	AA574382 nf46d02.s
29	399.4	31.0	429	30	AA224847	AA224847 nc33c12.s
30	391	30.3	439	28	AA088704	AA088704 z183f08.s
31	388	30.1	400	50	A1670721	A1670721 wc28d05.x
32	383	29.7	431	29	AA146775	AA146775 zo35g09.s
33	382.8	29.7	655	50	A1673653	A1673653 we77b03.x
34	381.8	29.6	435	29	AA149408	AA149408 z125f07.r
35	381.8	29.6	444	30	AA225484	AA225484 nc25a04.r
36	379.6	29.4	388	33	AA420727	AA420727 nc63b06.s
37	378.2	29.3	388	35	AA557820	AA557820 n163b10.s
38	376	29.2	388	49	A1620534	A1620534 t056h12.x
39	375	29.1	399	42	A1091832	A1091832 qe58d02.s
40	373	28.9	407	29	AA161042	AA161042 zo74g11.r
41	373	28.9	373	34	AA468378	AA468378 nc78e09.s
42	373	28.9	399	40	AA931185	AA931185 cm97b04.s
43	371.8	28.8	382	34	AA468030	AA468030 nc72f07.s
44	368.4	28.6	392	49	A1660655	A1660655 wf23c06.x
45	368	28.5	428	28	AA113128	AA113128 zm29d06.s

## ALIGNMENTS

RESULT 1  
A1660579 697 bp mRNA EST 10-MAY-1999  
LOCUS we68d03.x1 Soares.Dieckgraefe\_colon NHCD Homo sapiens clone  
DEFINITION IMAGE:2346245 3' similar to TR:060635 060635 TSPAN-1. [1] ;, mRNA  
sequence.  
ACCESSION A1660579  
NID 94764162

VERSION A1660579.1 GI:4764162  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 697)  
NCI-CCGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949547.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40UP from Glibco  
High quality sequence stop: 455.  
Location/Qualifiers  
1..697  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="between D11S1765 and UGB"  
/clone="IMAGE:2346245"  
/clone\_lib="Soares.Dieckgraefe.colon\_NHCD"  
/tissue\_type="colonic mucosa from 3 patients with Crohn's  
disease"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: colon; Vector: pYT73D-Pac (Pharmacia) with a  
modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st  
strand cDNA was primed with a Not I - Oligo(dT) primer [5'  
TGTTACCACTGCAAGTGGAGGCGCCGCTTTTCTTTTCTTTT 3'],  
double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pYT73 vector. Library  
went through one round of normalization. Tissue samples  
provided by Dr. Brian Dieckgraefe (Washington University,  
dieckgrbm.wustl.edu); colonic mucosa represents a range of  
disease involvement from moderate to severe Crohn's  
disease; samples include both perforating (fistulas) and  
non-perforating samples. Library constructed by Bento  
Soares and M. Fatima Bonaudo.  
BASE COUNT 168 a 182 c 170 g 175 t 2 others  
ORIGIN

Query Match 48.2%; Score 621.6; DB 49; Length 697;  
Best Local Similarity 95.8%; Pred. No. 5.1e-163;  
Matches 669; Conservative 0; Mismatches 26; Indels 3; Gaps 3;

QY 583 ACATACGAGATTGAGAGCTACCCCTACTTCAAGAGAAAGAGCTTTCCCATCTT 642  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 696 ACATCGGATTTTGAGATACCTACCTCAAGAGAGAGAGCTTTCCCATCTTCT 637  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 643 GTTGCAATGACAGGTACACACAGCAGCAATGAACCTGACACCAAGAGAGCTACG 702  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 636 GTTGCAATG-CAGCGTACACACAGCAGCAATGAACCGACCAAGAGAGCTACG 578  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 703 ACCAAAAGTAGAGGGTGGCTTCATCAGCTTTTGATGACATCCGAACTAATGCACTCA 762  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 577 A-CANNAAGTAGAGGGTGGCTTCATCAGCTTTTGATGACATCCGAACTAATGCACTCA 519  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 763 CCGTGGGTGTGTGGCAGCTGGAATTTGGGGCCCTCGAGCTGGCTGCATATTTGTGCCA 822  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 518 CCGTGGGTGTGTGGCAGCTGGAATTTGGGGCCCTCGAGCTGGCTGCATATTTGTGCCA 459  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 823 TGTATCTGTACTGCAATCTCAATTAAGTCACTTCTGCGCATCTACTGCTGCCACA 882  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 458 TGTATCTGTACTGCAATCTCAATTAAGTCACTTCTGCGCATCTACTGCTGCCACA 399  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 883 TGGGAACGTGGAAGAGCCACCTGGCAAGCAGCAGTGTGGGGAGGAGACAGATCTA 942

Db 398 TGGGAACGTGGAAGAGCCACCTGGCAAGCAGCAGTGTGGGGAGGAGACAGATCTA 339  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 943 ACAATGCTACTTGGGCGCAATATGAGACTTCCCTTTTGTCTCCAGACTTGGGCTAGATAG 1002  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 338 ACAATGCTACTTGGGCGCAATATGAGACTTCCCTTTTGTCTCCAGACTTGGGCTAGATAG 279  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 1003 GGACACCTCTTTTA-GCGATGGCTACTTCTTCCATTTGGTGGGTGATGGGAGG 1061  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 278 GGACACCTCTTTTATAGCCATGCTGCTACTTCTTCCATTTGGTGGGTGATGGGAGG 219  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 1062 GCATTCCAGAGCCCTTAAGTAGAGCCAGTTCTGTGCCATTCCTCCAGTCTATTAAACC 1121  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 218 GCATTCCAGAGCCCTTAAGTAGAGCCAGTTCTGTGCCATTCCTCCAGTCTATTAAACC 159  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 1122 TTGATATGCCCCCTAGGCGCTAGTGTGATCCAGAGTCTTACTGGGGATGAGAGAAAG 1181  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 158 TTGATATGCCCCCTAGGCGCTAGTGTGATCCAGAGTCTTACTGGGGATGAGAGAAAG 99  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 1182 CATTTATAGCCCTGGCATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 1241  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 98 CATTTATAGCCCTGGCATTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAGTAAG 39  
||||| ||||||| ||| ||||| ||||||| |||||||  
QY 1242 ACTTCAAAATGCATAAACCTGTACAAATGTAAAAAA 1279  
||||| ||||||| ||| ||||| ||||||| |||||||  
DB 38 ACTTCAAAATGCATAAACCTGTACAAATGTAAAAAA 1  
||||| ||||||| ||| ||||| ||||||| |||||||

RESULT 2  
AA100799/c 593 bp mRNA EST 23-DEC-1997  
LOCUS zm26d01.s1 Striatogene pancreas (#937208) Homo sapiens cDNA clone  
DEFINITION IMAGE:526753 3', mRNA sequence.  
ACCESSION AA100799  
NID 91647216  
VERSION AA100799.1 GI:1647216  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 593)  
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chispe, S.,  
Chispe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Plenge, C., Rifkin, L.,  
Roifling, T., Scheinberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,  
Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.  
and Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
genome Res. 6 (9), 807-828 (1996)  
97044478  
On May 8, 1995 this sequence version replaced gi:800260.  
COMMENT

TITLE  
JOURNAL  
MEDLINE  
GENE  
CONTACT  
WILSON  
WASHINGTON UNIVERSITY SCHOOL OF MEDICINE  
4444 FOREST PARK PARKWAY, BOX 8501, ST. LOUIS, MO 63108  
TEL: 314 286 1800  
FAX: 314 286 1810  
EMAIL: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
WARNING: There is evidence that suggests that the 384-well parent  
plate of this clone contains both human and mouse derived clones.  
Thus, the origin of this clone is uncertain. This caution should be  
kept in mind should you use this clone.

This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 353.  
Location/Qualifiers  
1..593  
/organism="Homo sapiens"  
/db\_xref="GDB:3918182"

```

/db_xref="taxon:9606"
/clone="IMAGE:526753"
/clone_1ib="Stratagene pancreas (#937208)"
/_db_host="SOLR cells (kanamycin resistant)"
/_note="Organ: pancreas; Vector: pBluescript SK-; Site_1
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pancreatic adenocarcinoma cell line. Average
insert size: 1.0 kb. Uni-ZAP XR Vector: -5' adaptor
sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5'
CTGACGCTTTTTTTTTTTTTTTT 3'"

```

Query Match	42.3%	Score 545	DB 28	Length 593
Best Local Similarity	97.8%	Pred	No. 1.1e-141	
Matches 570	Conservative	0	Mismatches 11	Indels 2
				Gaps 2

OY	693	AMGGCTACGACGACAAAAGAGAGGGGTGGCTCAACACCTTTGTATGACATCCGAAC	752
Db	582	AAAGGCCGCCACCAAAAAGTA-AAAGTGCCTTCAACACTTTNGTATGACATCCGAAC	524
OY	753	AATCAGATCACCGTGGGTGGTGTGG-CAGCTGGAATGGGGGCTCGAGCTGGTCCAT	811
Db	523	AATCAGATCACCGGNTGGTNTGGCAGGTGGAATGGGGGCTCGAGCTGGTCCAT	464
OY	812	GATTGTGTCCATGTATCTGTACTGCAATCTACAAATAGTCACCTTGTGCTTCGCACTA	871
Db	463	GATTGTGTCCATGTATCTGTACTGCAATCTACAAATAGTCACCTTGTGCTTCGCACTA	404
OY	872	CTGCTGCACATGGGAACCTGTGAAGAGGCCCTGGCAAGCAGAGTATGGGGGAGG	931
Db	403	CTGCTGCACATGGGAACCTGTGAAGAGGCCCTGGCAAGCAGAGTATGGGGGAGG	344
OY	932	GACAGATCTAACAAATGTCACCTTGGGCCGAAGATGAGACCTGCCCTTGTGCTCAACACTG	991
Db	343	GACAGATCTAACAAATGTCACCTTGGGCCGAAGATGAGACCTGCCCTTGTGCTCAACACTG	284
OY	992	GGGCTAAGTAGGGAACCACTCCTTTTAGCGATGCTTACTTTCCCTTCATGGTGGGTGA	1051
Db	283	GGGCTAAGTAGGGAACCACTCCTTTTAGCGATGCTTACTTTCCCTTCATGGTGGGTGA	224
OY	1052	TGGGTGGGGGCAATTCAGAGACCTCTAAGGTAGCCAGTCTGTGGCAATCCCCCACTC	1111
Db	223	TGGGTGGGGGCAATTCAGAGACCTCTAAGGTAGCCAGTCTGTGGCAATCCCCCACTC	164
OY	1112	TATTTAAACCTTGTATATGCCCCCTAGGGCCATAGTGTGATCCCAAGTCTACTAGGGGAT	1171
Db	163	TATTTAAACCTTGTATATGCCCCCTAGGGCCATAGTGTGATCCCAAGTCTACTAGGGGAT	104
OY	1172	GAGGAAAAGCATTTTATAGCCTGGGCAATAGTGAATAGCAGAGACCTGTGGTGTATG	1231
Db	103	GAGGAAAAGCATTTTATAGCCTGGGCAATAGTGAATAGCAGAGACCTGTGGTGTATG	44
OY	1232	TGTAGAAAGCACTTCAAAATGCAATTAACCGTGTACAATCTTAA 1274	
Db	43	TGTAGAAAGCACTTCAAAATGCAATTAACCGTGTACAATCTTAA 1	

RESULT	3
LOCUS	AI708181/c
DEFINITION	AI708181 557 bp mRNA EST 04-JUN-1999
LOCUS	AB991608.1 Barcode colon HPRB7 Homo sapiens cDNA clone IMAGE:2335911 3 similar to TR:060635 060635 TSPAN-1. [1] ;, mRNA sequence.
ACCESSION	AI708181
NID	G4997957
VERSION	AI708181.1 GI:4997957
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE	TITLE	JOURNAL	COMMENT
1 (bases 1 to 557)			
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kitzman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Mair, M., Martin, J., Moore, B., Schellenberg, K., Stepec, M., Tan, F., Thesling, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.	WashU-NCI human EST Project		
Unpublished (1997)			
On Dec 20, 1995 this sequence version replaced g1:1133478.			

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: [estewatson.wustl.edu](mailto:estewatson.wustl.edu)  
This clone is available royalty-free through LINL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Possible reversed clone: similarity on wrong strand  
Seq primer: -40UP from Gibco  
High quality sequence stop: 395.

```

FEATURES
SOURCE
Location/Qualifiers
1..557
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="1g"
/clone_1ID="IMAGE:2335911"
/clone_1ID="Barstead colon HPLRB7"
/sex="male"
/dev_stage="adult, age 25"
/lab_host="DH10B (phage resistant)"
/note="Organ: colon; Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACGATCTGAAGGGCGGCCGCCTTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors [5' AATTCACTAGTAAT 3' and 5' ATTACTAATG 3'], digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library constructed by Bob Barstead."
BASE COUNT
145 a      157 c      132 g      123 t
ORIGIN
```

Query Match	40.2%	Score 518;	DB 51;	Length 557;
Best Local Similarity	97.8%;	Orig. No.3,7e134;		
Matches 546;	Conservative 97;	Mismatches 10;	Indels 2;	Gaps 2
OY 717	GATTCTTCATACAGCTTTTGATATACATCCGAACTAATGACAGTCACCGCTGGGTGGTG	776		
Db 557	GGGTCTTCATACAGCTTTTGATATACATCCGAACTAATGACAGTCACCGCTGGGTGGTG	498		
OY 777	GCAGCTGGAATTTGGGGGCGCTCAGCTGGCTGCATGATTTGTCTCATGTATCTGACTGC	836		
Db 457	GCAGCTGGAATTTGGGGGCGCTCAGCTGGCTGCATGATTTGTCTCATGTATCTGACTGC	438		
OY 837	AATCTACATAAAGTCCACTTCTGCTCTGCCACTACTGCTGCCACATGGGAAGTGTGAAG	896		
Db 437	AATCTACATAAAGTCCACTTCTGCTCTGCCACTACTGCTGCCACATGGGAAGTGTGAAG	378		
OY 897	AGGCACCGCTGGCAACACAGAGATTTGGGGGGGAGGAGACAGATCTTAACAATCTCACTGG	956		
Db 377	AGGCACCGCTGGCAACACAGAGATTTGGGGGGGAGGAGACAGATCTTAACAATCTCACTGG	319		
OY 957	GCCAGAAATGACACTGCCCTTCTGCTGCACAGATTGGGCTAATAGGACCACTCCTTTT	1016		
Db 318	GCCAGAAATGACACTGCCCTTCTGCTGCACAGATTGGGCTAATAGGACCACTCCTTTT	259		
OY 1017	A-GGCATCCTGACTTTCCTTCATATGATGGGTGGATGGGTGGGGGCACTTCACAGAGCT	1075		
Db 258	AGGCATCCTGACTTTCCTTCATATGATGGGTGGATGGGTGGGGGCACTTCACAGAGCT	199		
OY 1076	CTAAGGTAGCCAGTTCTGTGGCCATTTCGCCAGTCTATAACCTTGATATGCCCGCT	1135		



QY	705	CAAAATAGAGGGTTCCTTCATACAGCTTTGTATGACATCGAACTAAATGACATCC	764
Db	1	CAAAATATGAGGGTTCCTTCATACAGCTTTGTATGACATCGAACTAAATGACATCC	60
QY	765	GTGGGTGGTGTGGCACTGGAATTTGGGGGCTTGAGCTGGCTGCATGATTTGTCCATG	824
Db	61	GTGGGTGGTGTGGCACTGGAATTTGGGGGCTTGAGCTGGCTGCATGATTTGTCCATG	120
QY	825	TATGTGTACATCTACAAATTAATGTCATTTTGCTCTGCGCATCTACGTCTGCCCATG	884
Db	121	TATGTGTACATCTACAAATTAATGTCATTTTGCTCTGCGCATCTACGTCTGCCCATG	180
QY	885	GGAAGCTGTGAAGAGGACCCCTGGCAAGACAGACTGATTTGGGGAGGGAGCAGATCTAAC	944
Db	181	GGAAGCTGTGAAGAGGACCCCTGGCAAGACAGACTGATTTGGGGAGGGAGCAGATCTAAC	240
QY	945	AATGTCACTTGGGCGAGATGAGACCTGCTCTTCTGCTCCAGACTTGGGGCTAGATAGGG	1004
Db	241	AATGTCACTTGGGCGAGATGAGACCTGCTCTCTCCAGACTTGGGGCTAGATAGGG	300
QY	1005	ACCACTCCTTTTAGCGATGCGCTGACTTCCTTCATTTGGGGGTGGATGGGTGGGGGCA	1064
Db	301	ACCACTCCTTTTAGCGATGCGCTGACTTCCTTCATTTGGGGGTGGATGGGTGGGGGCA	360
QY	1065	TTCCAGAGGCTTAAAGTAGCCAGATTCGTGTGGCCATTCGCCCAAGTCTTAAACCTTGG	1124
Db	361	TTCCAGAGGCTTAAAGTAGCCAGATTCGTGTGGCCATTCGCCCAAGTCTTAAACCTTGG	420
QY	1125	ATATGCCCCCTTAGGCCCTAGTGGTGATGCCAGTGTCTACTGSGGGATAGAGAAAGCAT	1184
Db	421	ATATGCCCCCTTAGGCCCTAGTGGTGAT -CCAGTGTCTACTGSGGGATAGAGAAAG -GCAT	478
QY	1185	TTTATAGCTGTGGGATAGTGAATCAGACAGACCTCTGGGTGGATGTATAGAGGCACT	1244
Db	479	TTTATAGCT -GGCATTAAGTG -AATCAGCAGAG -CTCTGGGTGGATGTATAGAG -CAC	533
QY	1245	TCAAAATGACATAAAGCTGTACATGTTAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTT	1289

```

Insert length: 1109      Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 390.
Location/Qualifiers
1. 479
/organism="Homo sapiens"
/db_xref="GDB:5923237"
/db_xref="taxon:9606"
/clone="IMAGE:745907"
/clone_id="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/notice="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected, histologically normal
prostate epithelial cells. Double-stranded cDNA was
ligated to EcoRI adaptors. 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the upc-Cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directional! cloned. This library was
constructed by David Kirizman."
BASE COUNT      122 a      130 c      119 g      108 t
ORIGIN
Query Match      37.0%; Score 477.4; DB 33; Length 479;
Best Local Similarity 99.8%; Prid. No. 7.4e-123;
Matches 478; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY      794      CCTCAGCTGGTGCATGATGTGTCCATGATCTGTACTCAATCTACATAAAGTCCA      853
Db      479      CCTCAGCTGGTGCATGATGTGTCCATGATCTGTACTCAATCTACATAAAGTCCA      420
OY      854      CTTGTGCTGTGCACATCTGCTGCACATGGGAATGTGAAGAGCACCTGGCAAGCA      913

```

```

Db 419 CTTCTGCGCTGCGCACTACTGTCGCCACATGGAGACTGTGAAGAGCACCCTGGCAAGCA 360
Oy 914 GCAGTATTTGGGGAGGGGACAGATCTAACAAATGTACTTGGGCCAGATGAGACCTGCC 973
Db 359 GCAGTATTTGGGGAGGGGACAGATCTAACAAATGTACTTGGGCCAGATGAGACCTGCC 300
Oy 974 CTTTCTACTCAGACTGGGGCTAGATAGGAGCACCTCTTTTATAGCATGCTGACTTTC 1033
Db 299 CTTTCTACTCAGACTGGGGCTAGATAGGAGCACCTCTTTTATAGCATGCTGACTTTC 240
Oy 1034 CTTTCATTTGGTGGTGGATGGTGGGGGCAATTCAGAGCCTTAAGATAGCAATTCG 1093
Db 239 CTTTCATTTGGTGGTGGATGGTGGGGGCAATTCAGAGCCTTAAGATAGCAATTCG 180
Oy 1094 TTGGCCCTTCCCGACAGCTTTAAACCCCTGATATGCCCCCTAGCCCTAGTGGATGCC 1153
Db 179 TTGGCCCTTCCCGACAGCTTTAAACCCCTGATATGCCCCCTAGTGGATGCC 120
Oy 1154 AGTCTCTACTGGGGGATGAGAGAAAGCATTTTATAGCCTGGGCATAGTGAATGAGC 1213
Db 119 AGTCTCTACTGGGGGATGAGAGAAAGCATTTTATAGCCTGGGCATAGTGAATGAGC 60
Oy 1214 AGAGCCTCTGGGTGGATGTGTAGAGGCACTTCAAAAATGCAATAAAGCTTTACATGTT 1272
Db 59 AGAGCCTCTGGGTGGATGTGTAGAGGCACTTCAAAAATGCAATAAAGCTTTACATGTT 1

RESULT 7
AA468417 505 bp mRNA 14-AUG-1997
LOCUS nc78609.r1 NCI_CGAP_Pr2 Homo sapiens cDNA clone IMAGE:78496, mRNA
DEFINITION sequence.
ACCESSION AA468417
NID 92194951
VERSION AA468417.1 GI:2194951
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 505)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1292183.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,
M.D., Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Kitzman, Ph.D.
DNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 703 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amerisham
High quality sequence stop: 442.
Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:78496"
/clone_1ib="NCI_CGAP_Pr2"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/Note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of

```

```

BASE COUNT 109 a 123 c 144 g 129 t
ORIGIN
Query Match 36 1%; Score 464.8; DB 34; Length 505;
Best Local Similarity 98.2%; Pred. No. 2,5e-119;
Matches 491; Conservative 0; Mismatches 7; Indels 2; Gaps 2;

DNAME-treated, total cellular RNA obtained from
5,000-10,000 microdissected preneoplastic cells
histologically-determined to be prostatic intraepithelial
neoplasia 2 (PIN2) cells. Double-stranded cDNA was
ligated to EcoRI adaptors, 5 cycles of PCR applied to the
cDNA with an adaptor-specific primer, and the resulting
PCR product subcloned into pAMP10 by the uDG-cloning
method (Life Technologies). Average insert size is 600
bp. NOTE: Not directionally cloned. This library was
constructed by David Kitzman.

Oy 701 CGACCAAAAGTAGAGG-GTTGCTCAATCAGCTTTGTATGACATCCGACTAATGACG 759
Db 6 CGACCAAAAGTAGAGGATTTATTAATCAGCTTTGTATGACATCCGACTAATGACG 65
Oy 760 TCACCGTGGGTGGTGTGGCAGCTGGAA-TTGGGGGCTTCGAGCTGGCTGCATGATGTG 818
Db 66 TCACCGTGGGTGGTGTGGCAGCTGGAA-TTGGGGGCTTCGAGCTGGCTGCATGATGTG 125
Oy 819 TCCATGTATCTGTACTGCAATCTACAAATAGTCCATTTGGCTCTGCCACTACTGCTGC 878
Db 126 TCCATGTATCTGTACTGCAATCTACAAATAGTCCATTTGGCTCTGCCACTACTGCTGC 185
Oy 879 CACATGGGAAGTGTGAAGAGGACACCTGGCAAGCAGCATGATTTGGGGAGGGAGCAGGA 938
Db 186 CACATGGGAAGTGTGAAGAGGACACCTGGCAAGCAGCATGATTTGGGGAGGGAGCAGGA 245
Oy 939 TCTAACATGTCACTTGGGCCAGATGACCTGCCCCCTTCTGCTCCAGACTTGGGGCTAG 998
Db 246 TCTAACATGTCACTTGGGCCAGATGACCTGCCCCCTTCTGCTCCAGACTTGGGGCTAG 305
Oy 999 ATAGGGACACACCTTTTAGCCATGCTGCTTCCATTTGGTGGGTGGATGGTGG 1058
Db 306 ATAGGGACACACCTTTTAGCCATGCTGCTTCCATTTGGTGGGTGGATGGTGG 365
Oy 1059 GGGGCAATCCAGAGCCTCAAGGTAGCAGTCTGTGGCCATCCCGCAGCTATTAA 1118
Db 366 GGGGCAATCCAGAGCCTCAAGGTAGCAGTCTGTGGCCATCCCGCAGCTATTAA 425
Oy 1119 CCCTTATATGCCCCCTAGGCTAGTGTGATCCAGTCTCTACTTGGGGGATGAGAGAA 1178
Db 426 CCCTTATATGCCCCCTAGGCTAGTGTGATCCAGTCTCTACTTGGGGGATGAGAGAA 485
Oy 1179 AGGCATTTATAGCCTGGGC 1198
Db 486 AGGCATTTATAGCCTGGGC 505

RESULT 8
AA593864 470 bp mRNA 25-SEP-1997
LOCUS aa593864/c n119f08.s1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1084359 3',
DEFINITION mRNA sequence.
ACCESSION AA593864
NID 92408542
VERSION AA593864.1 GI:2408542
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 470)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)

```



COMMENT On May 5, 1995 this sequence version replaced gl:797843.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 825 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 435.

## FEATURES

source

1. 470  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1084359"  
/clone\_lib="NCI\_CGAP\_Col2"  
/sex="mixed"  
/tissue="colon tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: colon; Vector: Bluescript SK-; Site.1:  
EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'  
GAATTCGCGACGAG 3' 3' adaptor sequence: 5'  
CTCAGCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."  
BASE COUNT 121 a 127 c 114 g 107 t 1 others

ORIGIN

Query Match 35.5%; Score 458; DB 35; Length 470;  
Best Local Similarity 99.8%; Pred. No. 1.9e-117;  
Matches 469; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

804 GCTGCCATGATGTGTCTCATGTATCTGTACTGCATCTACAAATAGTCACCTTCTGCTC 863

470 GCTGCCATGATGTGTCTCATGTATCTGTACTGCATCTACAAATAGTCACCTTCTGCTC 411

864 TGGCAGTACGCTGCCATGAGGAAGCTGTGAAGAGCAACCTGGCAGACGACGATGATG 923

410 TGGCAGTACGCTGCCATGAGGAAGCTGTGAAGAGCAACCTGGCAGACGACGATGATG 351

924 GGGGAGGAGACAGATCTAACAATGTCACTTGGCCAGAAATGACCTGCTTCTGCTC 983

350 GGGGAGGAGACAGATCTAACAATGTCACTTGGCCAGAAATGACCTGCTTCTGCTC 291

984 CAGACTTGGGGCTAGATAGGAGCAGCTCTTTT AGCGATGCTGACTTCTTCTGATG 1042

290 CAGACTTGGGGCTAGATAGGAGCAGCTCTTTT NAGGATGCCGACTTCTTCTGATG 231

1043 GTGGGTGATAGGGGGGGGGGCGATCCAGAGCCTTAAGATAGGACGATCTTCTGCTC 1102

230 GTGGGTGATAGGGGGGGGGGCGATCCAGAGCCTTAAGATAGGACGATCTTCTGCTC 171

1103 CCCCAGCTATTTAAACCTTGATATGCCCTAGAGGCTAGTGTGATCCAGATGCTCTA 1162

170 CCCCAGCTATTTAAACCTTGATATGCCCTAGAGGCTAGTGTGATCCAGATGCTCTA 111

1163 CTGGGGGATAGAGAAAGCATTTTATAGCTGGGCAATAGTAAATCAGACGAGCCTCT 1222

110 CTGGGGGATAGAGAAAGCATTTTATAGCTGGGCAATAGTAAATCAGACGAGCCTCT 51

1223 GGGGTGATGTGTAGAGGCACTTCAAAATGATTAACCTGTATAAATGTT 1272

50 GGGGTGATGTGTAGAGGCACTTCAAAATGATTAACCTGTATAAATGTT 1

## FEATURES

source

AA037886/c  
LOCUS 488 bp mRNA EST 01-FEB-1997  
DEFINITION 2f04h10.s1 Soares fetal\_heart\_NbH919W Homo sapiens cDNA clone  
IMAGE:376003 3', mRNA sequence.  
ACCESSION  
AA037886  
NID 91513022  
VERSION AA037886.1 GI:1513022  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
Hiller, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiappelli, B.,  
Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,  
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,  
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Ritzke, L.,  
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierly-Meg, J.,  
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,  
Marra, M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478  
COMMENT  
On Sep 21, 1992 this sequence version replaced gl:279014.

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
This clone is available royalty-free through LLNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1834 Std Error: 0.00  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 434.

Location/Qualifiers

1. 488  
/organism="Homo sapiens"  
/db\_xref="GDB:1284259"  
/db\_xref="taxon:9606"  
/map="19p12-p13.1"  
/clone="IMAGE:376003"  
/clone\_lib="Soares\_fetal\_heart\_NbH919W"  
/sex="unknown"  
/dev\_stage="19 weeks"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: heart; Vector: pRT73D (Pharmacia) with a  
modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st  
strand cDNA was primed with a Not I - oligo(dt) primer (5'  
TGTTCACAACTGTAAGTGGAGCGCGCGCATCTTTTCTTTTCTTTT 3'),  
double-stranded cDNA was size selected, ligated to Eco RI  
adapters (Pharmacia), digested with Not I and cloned into  
the Not I and Eco RI sites of a modified pRT73 vector  
(Pharmacia). Library went through one round of  
normalization to a Cot = 5. Library constructed by  
M.Fatima Bonaldo. This library was constructed from the  
same fetus as the fetal lung library, Soares fetal lung  
NBH19W."

BASE COUNT 123 a 134 c 121 g 108 t 2 others

ORIGIN

Query Match 35.4%; Score 456.8; DB 27; Length 488;  
Best Local Similarity 99.0%; Pred. No. 4.2e-117;  
Matches 480; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

790 GGGGCTCGAGCTGGCGCCATGATGTGTCTCATGTATCTGTACTG-CAATCTACAAATA 848

488 GGGGCTCGAGCTGGCGCCATGATGTGTCTCATGTATCTGTACTG-CAATCTACAAATA 429

849 GTCCACTTGTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 908

RESULT 9

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="672C09:17:17p11.2:21q"
/clone="IMAGE:1387451"
/clone_1lb="Soares.parathyroid_tumor_NDHPA"
/tissue_type="parathyroid tumor"

```

SOURCE	ORGANISM
human.	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
	Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE	1 (bases 1 to 450)
AUTHORS	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 17, 1998 this sequence version replaced g1:1900763.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Strategene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html

Insert Length: 1215 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 379.

FEATURES  
SOURCE

1.450  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="159433"  
/clone\_lib="NCI\_CGAP\_Col2"  
/sex="mixed"  
/tissue\_type="colon tumor"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Organ: colon; Vector: Bluescript SK-; Site\_1:  
EcORI; Site\_2: XhoI; Cloned unidirectionally. Primer:  
Oligo dt. Pooled colon tumors. 5' adaptor sequence: 5'  
GAATGGCAGAG 3', 3' adaptor sequence: 5'  
CTCGCTTTTCTTTTCTTTTCTTTT 3' Average insert size: 1.2 kb."  
BASE COUNT 114 a 123 c 109 g 104 t  
ORIGIN

Query Match 34.0%; Score 438; DB 42; Length 450;

Best Local Similarity 99.8%; Pred. No. 7.1e-112;  
Matches 449; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

824 GATCTGTACTGCAATCTACATTAAGTCACCTTCTGCTGCACTACTGTCACAT 883  
|||||  
450 GATCTGTACTGCAATCTACATTAAGTCACCTTCTGCTGCACTACTGTCACAT 391  
|||||  
884 GGAAGCTGTGAAGAGCACCCTGGCAAGCAGAGTGTGGGGAGGGAGCAGATCTAA 943  
|||||  
390 GGAAGCTGTGAAGAGCACCCTGGCAAGCAGAGTGTGGGGAGGGAGCAGATCTAA 331  
|||||  
944 CATGTCTACTTGGGGCCGAGATGAGACCTGCTTCTCTCCAGACTTGGGGCTAGATAGG 1003  
|||||  
330 CAATGTCTACTTGGGGCCGAGATGAGACCTGCTTCTCTCCAGACTTGGGGCTAGATAGG 271  
|||||  
1004 GACCACCTCTTTTA-GCGATGCTGACTTCTCTCCATGTTGGGATGATGGGAGG 1062  
|||||  
270 GACCACCTCTTTTAAGGATGCTGACTTCTCTCCATGTTGGGATGATGGGAGG 211  
|||||  
1063 CATTCAGAGCCTCTTAAGTACAGCAGTGTGTTGCCCATTTCCCCAGCTATTAAACCT 1122  
|||||  
210 CATTCAGAGCCTCTTAAGTACAGCAGTGTGTTGCCCATTTCCCCAGCTATTAAACCT 151  
|||||  
1123 TGATATGCCCCCAGAGCCTAGTGTATCCAGTCTCTACTGGGGATAGAGAAAGC 1182  
|||||  
150 TGATATGCCCCCAGAGCCTAGTGTATCCAGTCTCTACTGGGGATAGAGAAAGC 91  
|||||  
1183 ATTATATAGCCTGGGCAATAGTAAATCAGACAGACCTTGGGATGATGTAGAAAGCA 1242  
|||||  
90 ATTATATAGCCTGGGCAATAGTAAATCAGACAGACCTTGGGATGATGTAGAAAGCA 31  
|||||  
1243 CTTCAAATGCAATAAACCTGTACAAATGTT 1272  
|||||  
30 CTTCAAATGCAATAAACCTGTACAAATGTT 1

RESULT 12  
AA659579

LOCUS AA659579 453 bp mRNA EST 05-NOV-1997  
DEFINITION nt63b06.s1 NCI\_CGAP\_pr3 Homo sapiens CDNA clone IMAGE:1203155  
ACCESSION AA659579  
NID 92595733  
VERSION AA659579.1 GI:2595733  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 8, 1995 this sequence version replaced g1:801529.

FEATURES  
SOURCE

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagut,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbp/image/image.html  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 436.  
Location/Qualifiers  
1.453  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="1203155"  
/clone\_lib="NCI\_CGAP\_Pr3"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dt)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically-determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI  
adaptors, 5 cycles of PCR applied to the cDNA with an  
adaptor-specific primer, and the resulting PCR product  
subcloned into PAMP10 by the UDS-cloning method (Life  
Technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Kitzman."

BASE COUNT 93 a 113 c 133 g 114 t  
ORIGIN

Query Match 33.9%; Score 437.6; DB 36; Length 453;

Best Local Similarity 98.0%; Pred. No. 9.2e-112;  
Matches 443; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

741 GACATCGAAGTAACTACAGTACCGTGGGTGTGTGGCAGTGAATGGGGCTCGAG 800  
|||||  
2 GTGACCGAAGTAACTACAGTACCGTGGGTGTGTGGCAGTGAATGGGGCTCGAG 61  
|||||  
801 CTGGCTGCATGATGTGTTCATGATCTGTACTGCAATCTACAAATAGTCCACTTCTGC 860  
|||||  
62 CTGGCTGCATGATGTGTTCATGATCTGTACTGCAATCTACAAATAGTCCACTTCTGC 121  
|||||

Qy	861	CTCTCCACACTACTGCTGGCACATGGGAACCTGGAAGAGGACACCTCGGCAAGGAGAGTA	920
Db	122	CTCTGCACTACTGCTGTCACATGGGAACCTGGAAGAGGACACCTCGGCAAGGAGAGTA	181
Qy	921	TTGGGGGAGGGGACAGACTCTAACAACTGCTACTTGGGCCAGATGAGACTGCGCTTTCTG	980
Db	182	TTGGGGGAGGGGACAGACTCTAACAACTGCTACTTGGGCCAGATGAGACTGCGCTTTCTG	241
Qy	981	CTCCAGACTTTGGGGCTAGATAGGAGCACTCCTTTTAGCGATGCTGACTTTCCTTCAT	1040
Db	242	CTCCAGACTTTGGGGCTAGATAGGAGCACTCCTTTTAGCGATGCTGACTTTCCTTCAT	301
Qy	1041	TGCTGGGTGGGATGGGTGGGGGGGCAATTCACAGACCTCTTAAGTAAAGCAAGTTCTGTTGCCA	1100
Db	302	TGCTGGGTGGGATGGGTGGGGGGGCAATTCACAGACCTCTTAAGTAAAGCAAGTTCTGTTGCCA	361
Qy	1101	TTCCCCACAGTCTATTAAACCCTTGATATGCCCCCTAGGCGCTAGTGGTGAATCCAGATGCTC	1160
Db	362	TTCCCCACAGTCTATTAAACCCTTGATATGCCCCCTAGGCGCTAGTGGTGAATCCAGATGCTC	421
Qy	1161	TACTGGGGGATGAGAAAGGCAATTTTATAGC	1192
Db	422	TACTGGGGGATGAGAAAGGCAATTTTATAGC	453

RESULT	13				
AA534171/c					
LOCUS	AA534171	486 bp	mRNA	EST	21-AUG-1997
DEFINITION	U70b07.s1	NCI-CGAP_P10	Homo sapiens	CDNA	clone IMAGE:997813, mRNA
ACCSSION	AA534171				
NID	AA534171				
VERSION	92278187				
	AA534171.1	GI:2278187			

ORGANISM	SOURCE
<i>Homo sapiens</i>	human.

REFERENCE  
1 (bases 1 to 496)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia  
Eutheria; Primates; Catarrhini; Homidae; Homo.

**AUTHORS** NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
**TITLE** National Cancer Institute, Cancer Genome Anatomy Project (CGAP)  
**JOURNAL** Tumor Gene Index  
**COMMENT** Unpublished (1997)  
 On Nov 29, 1993 this sequence version replaced g1:637695.

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Kitzman, Ph.D.  
cDNA Library Arrayed by: Greg Iannoni, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/BLND at:  
[www-bio.linnl.gov/bbrip/image/image.html](http://www-bio.linnl.gov/bbrip/image/image.html)

```

Insert Length: 568   Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 386.
      Location/Qualifiers
source          1..496

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:997813"
/clone.lib="NCI-CGAP_Pri0"
/sex="male"
/tissue.type="Invasive prostate tumor"
/lab_host="DH10B"
/notes="Organ: Prostate; Vector: pAMP10; mRNA made from
invasive prostate tumor, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel,
average insert size 600 bp. Library made by D. Kitzman

```

BASE COUNT	121 a	131 c	120 g	124 t
ORIGIN				
Query Match	33.7%			
Best Local Similarity	99.1%			
Matches 447; Conservative	0; Pred. No. 8.5e-11; DB 34; Length 496; indels 1; Gaps 1			

QY	840	CTACAAATTAAGCCACTCTCTCCCTGCGACATACGTGTGCACATGGGAACGTGTGAAGAG	899
Db	468	CTACATTAATGATCCACTCTCTGCTCTGTGCACATCTGCTGCCACATGGGAACGTGTGAAGAG	409
QY	900	CACCTTGCGAAGCAGCAGATGTTGGGGGAGGGAGACAGATCTAACATGTCACTTGGGCC	959
Db	408	CACCTTGCGAAGCAGCAGATGTTGGGGGAGGGAGACAGATCTAACATGTCACTTGGGCC	349
QY	960	AGAATGACCCGCCCCCTTCTGTGCTCCAGACTTGGGGCTAATAGGAGACCTCCTTTA-G	1018
Db	348	AGAAATGACCCGCCCCCTTCTGTGCTCCAGACTTGGGGCTAATAGGAGACCTCCTTTTAGG	289
QY	1019	CGATGCCGTACTTCTCTTCATTTGATGGGAGATGGGTGGGGGCAATTCAGAGCCTCTA	1078
Db	288	CGATGCCGTAATTCCTTCATTTGATGGGAGATGGGTGGGGGCAATTCAGAGCCTCTA	229
QY	1079	AGGTAGCCAGTTCTGTGGCCCAATCCCCAGCTATTTAAACCCCTGATATGCCCCCTAGG	1138
Db	228	AGGTAGCCAGTTCTGTGGCCCAATCCCCAGCTATTTAAACCCCTGATATGCCCCCTAGG	169
QY	1139	CCTAGTGGTATCCCACTGCTCTACTAGGGGGAGTGAAGAAAGCAATTTATATGCTTG6GC	1199
Db	168	CCTAGTGGTATCCCACTGCTCTACTAGGGGGAGTGAAGAAAGCAATTTATATGCTTG6GC	109
QY	1199	ATAAGTGAATCAGCAGAGCCCTGTGGGTGATGTGTAGAAGGCACCTCAAAATGCATAAA	1258
Db	108	ATAAGTGAATCAGCAGAGCCCTGTGGGTGATGTGTAGAAGGCACCTCAAAATGCATAAA	49
QY	1259	CCTGTACAATGTAAAAAAAAAAAAAAAAAAAA 1289	
Db	48	CCTGTACAATGTGAAAAAAAAAAAAAAAAAAAA 18	

RESULT	14
AA861869/c	
LOCUS	AA861869
DEFINITION	456 bp mRNA
ACCESSION	U04759
NID	U04759
VERSION	U04759
KEYWORDS	U04759
SOURCE	human
ORGANISM	Homo sapiens
REFERENCE	U04759
AUTHORS	U04759
TITLE	U04759
JOURNAL	U04759
COMMENT	U04759

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Felima Boraldo, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/TLN! at: [www-bio.linnl.gov/bdnp/image/image.html](http://www-bio.linnl.gov/bdnp/image/image.html)



Db 140 CTAGGCTTACTGGTGATGCCAGTCTCTACTGGGGATGAGAGAAAGCATTATTATAGCC 81  
QY 1194 TGGCATTAATGAATCAGAGAGCCTCTGGGTGATGTGTAGAGGCACTTCAAAATGC 1253  
Db 80 TGGCATTAATGAATCAGAGAGCCTCTGGGTGATGTGTAGAGGCACTTCAAAATGC 21  
QY 1254 ATAACTCTGTACATGTTA 1273  
Db 20 ATAACTCTGTACATGTTA 1

Search completed: September 25, 1999, 12:05:02  
Job time: 8106 sec



|||||  
86 ysa:n 87

seq\_name: A\_Geneseq\_36:P60096

seq\_documentation\_block:

ID P60096 standard; Protein: 255 AA.

AC P60096;

DT 27-JUN-1991 (first entry)

DE Sequence of protein having human cholinesterase (ChE)

DE activity in clone FPCHE12.

DE Organophosphorus poisoning; therapy: prophylaxis; diagnosis;

KW Organophosphorus poisoning; therapy: prophylaxis; diagnosis;

KW pseudocholinesterase deficiency.

OS Homo sapiens.

PN EP-206200-A.

PD 30-DEC-1986.

PF 16-JUN-1986; 108189.

PR 18-JUN-1985; IL-075553.

PA (YEDA ) YEDA RES & DEV CO LTD.

PI Soreq H;

DR WPI: 86-340581/52.

DR N-PSDB; N60110.

PT Human cholinesterase-type proteins - produced by host cell contg.

PT recombinant vector which contains DNA fragment coding for the

PT enzyme

PS Claim 17: Page 36; 44pp; English.

CC Human cholinesterase-type proteins can be used for the prophylaxis

CC or for the treatment of the effects of organophosphorus cpd.

CC poisoning. Other uses include clinical detection of

CC pseudocholinesterase deficiencies; elucidation of the active site

CC topography and the AA sequence of AChE for the development of rapid

CC simple clinical methods to detect poisoning or disease-related

CC changes in ChEs, and for detecting mutations in ChE genes.

SO Sequence 255 AA;

Alignment\_scores: Quality: 61.50 Length: 62 Gaps: 5 Ratio: 1.809 Percent Similarity: 54.839 Percent Identity: 29.032

Alignment\_block: US-09-030-606-115 x P60096 ..

Align seg 1/1 to: P60096 from: 1 to: 255

```

      8 CTCCTCCCTCTCTGAATTAATTCCTTCACTGCAATTCGACAGATT 57
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      1 ILeSerProLysTyTyrMetIlePheThrProCysLysLeu..... 14
      58 ACACATTCACATGATGATATTTGTTGCAAAAAAAGAGTCTT 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      15 .....CysHisLeuCysCysArgGluSerGluIleAsn 26
      108 TGGTTT...AAATTAATCT.....TGGTTTGTG 130
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      26 eHisSerLysValThrIleIleCysIleArgPheLeuPheThrPheLeu 42
      131 AATCCATCTGCTTTTCCCATTCGAACTAGTCAT 166
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      43 ...LeuLeuCysMetLeu...IleGlyLysSerHis 52

```

seq\_name: A\_Geneseq\_36:R37441

seq\_documentation\_block:

ID R37441 standard; Protein: 255 AA.

AC R37441;

DT 06-OCT-1993 (first entry)

DE Partial human pseudocholinesterase sequence encoded by FPCHE12.

KW butyrylcholinesterase; acetylcholine acylhydrolase; EC 3.1.1.8; psi-ChE;

KW pseudo-ChE; neurotransmitter; organophosphorus insecticide; OP-poison;

KW antidote.

OS Homo sapiens.

```

FH Key Location/Qualifiers
FT peptide 30..49
FT misc_difference 101..110
FT /note- "putative leader peptide"
FT /note- "shown in the specification as being encoded
FT by the sequence TCT CTG ACC AGG TGG TCT GAT
FT ATT TGG ACT which should correspond to the
FT amino acid sequence SLTRMSDIWT"
FT misc_difference 111..120
FT /note- "shown in the specification as being encoded
FT by the sequence GCC ACA AAA TAT GCA AAT TCT
FT TGC TGT CAG which should correspond to the
FT amino acid sequence ATRVANSOCO"
FT active_site 251
FT /note- "active site Serine"
FT PN US5215909-A.
FT PD 01-JUN-1993.
FT PF 18-JUN-1986; 875737.
FT PR 18-JUN-1986; US-875737.
FT PR 21-AUG-1987; US-087724.
FT PR 15-AUG-1990; US-572911.
FT PA (YEDA ) YEDA RES & DEV CO LTD.
FT PI Soreq H;
FT DR WPI: 93-188509/23.
FT DR N-PSDB; Q42495.
FT PT Recombinant human gene encoding human pseudo-cholinesterase -
FT used to treat organo-phosphorus poisoning
FT Claim 3: Columns 40 and 41: 34pp; English.
CC A cDNA library prepared from foetal brain mRNA was screened with
CC degenerate probe pools based on the organophosphorus binding site of
CC cholinesterases. A 764 nucleotide insert (designated FPCHE12) was
CC isolated from one positive clone and sequenced. This insert,
CC containing an ORF large enough to code for about half the subunit
CC size of human cholinesterase, was used as a probe to obtain the full-
CC length pseudocholinesterase sequence. See also Q42496.
SO Sequence 255 AA;

```

Alignment\_scores: Quality: 61.50 Length: 62 Gaps: 5 Ratio: 1.809 Percent Similarity: 54.839 Percent Identity: 29.032

Alignment\_block: US-09-030-606-115 x R37441 ..

Align seg 1/1 to: R37441 from: 1 to: 255

```

      8 CTCCTCCCTCTCTGAATTAATTCCTTCACTGCAATTCGACAGATT 57
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      1 ILeSerProLysTyTyrMetIlePheThrProCysLysLeu..... 14
      58 ACACATTCACATGATGATATTTGTTGCAAAAAAAGAGTCTT 107
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      15 .....CysHisLeuCysCysArgGluSerGluIleAsn 26
      108 TGGTTT...AAATTAATCT.....TGGTTTGTG 130
      ::|||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      26 eHisSerLysValThrIleIleCysIleArgPheLeuPheThrPheLeu 42
      131 AATCCATCTGCTTTTCCCATTCGAACTAGTCAT 166
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
      43 ...LeuLeuCysMetLeu...IleGlyLysSerHis 52

```

seq\_name: A\_Geneseq\_36:P60097

seq\_documentation\_block:

ID P60097 standard; Protein: 635 AA.

AC P60097;

DT 27-JUN-1991 (first entry)

DE Sequence of protein having human cholinesterase (ChE)

DE activity.

KW Organophosphorus poisoning; therapy: prophylaxis; diagnosis;



KW pseudocholesterase deficiency.  
 OS Homo sapiens.  
 PN EP-206200-A.  
 PD 30-DEC-1986.  
 PF 16-JUN-1986; 108189.  
 PR 18-JUN-1985; IL-075553.  
 PA (YEDA ) YEDA RES & DEV CO LTD.  
 PI Scoreq H;  
 DR MPI: 86-340581/52.  
 DR N-PSDB: N60111.  
 PT Human cholinesterase-type proteins - produced by host cell contg.  
 PT recombinant vector which contains DNA fragment coding for the  
 PT enzyme  
 PS Claim 17; Page 38-40; 44pp; English.  
 CC Human cholinesterase-type proteins can be used for the prophylaxis  
 CC or for the treatment of the effects of organophosphorous cpd.  
 CC poisoning. Other uses include clinical detection of  
 CC pseudocholesterase deficiencies; elucidation of the active site  
 CC topography and the AA sequence of AChE for the development of rapid  
 CC simple clinical methods to detect poisoning or disease-related  
 CC changes in ChEs, and for detecting mutations in ChE genes.  
 SO Sequence 635 AA;

Alignment\_scores:  
 Quality: 60.50 Length: 62  
 Ratio: 1.779 Gaps: 5  
 Percent Similarity: 54.839 Percent Identity: 29.032

Alignment\_block:  
 US-09-030-606-115 x P60097 ..

Align seg 1/1 to: P60097 from: 1 to: 635

```

      8 CTCCTCCCTCCTGATTAATTCATTGCAATTCGCAAGGANT 57
      ::::::::::: ::::::::::: :::::::::::
     10 IIESEPRDASNTYTYTMeIlePheHnProCysLysVal..... 23
     58 ACACATTCTACTGTGATGTATGTGTGCAAAAAAAGTGTCT 107
     24 .....CysHISLeuCyScysArgGluSerGluIleAsnM 35
     108 TGTGT...AAATTACT.....TGGTTTGTG 130
      ::::::::::: ::::::::::: :::::::::::
     35 eCHISerLysValThrIleIleCysIleArgPheLeuPheTrpPheVal 51
     131 AATTCATCTTGCTTTTCCCATTCGCACTAGTCAT 166
     52 ...LeuLeuCyScysMetLeu...IleGlyLysSerHis 61

```

seq\_name: A\_Geneseq\_36:R41509

seq\_documentation\_block:  
 ID R41509 standard; Protein; 635 AA.  
 AC R41509;  
 DT 06-OCT-1993 (first entry)  
 DE Full-length foetal human pseudocholesterase.  
 KW butyrylcholinesterase; acetylcholine acylhydrolase; EC3.1.1.8; psi-ChE;  
 KW pseudo-ChE; neurotransmitter; organophosphorus insecticide; OP-poison;  
 KW antidote.  
 OS Homo sapiens.  
 FH Key  
 FH peptide  
 FT Location/Qualifiers  
 FT 35..58  
 FT /note= "putative leader peptide"  
 FT 79..81  
 FT /note= "potential N-glycosylation site"  
 FT 168..170  
 FT /note= "potential N-glycosylation site"  
 FT 303..305  
 FT /note= "potential N-glycosylation site"  
 FT 318..320  
 FT /note= "potential N-glycosylation site"  
 FT modified\_site  
 FT 403..405

```

FT modified_site /note= "potential N-glycosylation site"
FT 506...508
FT /note= "potential N-glycosylation site"
FT modified_site 543...545
FT /note= "potential N-glycosylation site"
FT active_site 260
FT /note= "active site serine"
FT US5215909-A.
PN 01-JUN-1993.
PD 18-JUN-1986; 875737.
PF 18-JUN-1986; US-875737.
PR 18-JUN-1986; US-087724.
PR 21-AUG-1987; US-087724.
PR 15-AUG-1990; US-572911.
PA (YEDA ) YEDA RES & DEV CO LTD.
PI Scoreq H;
DR MPI: 93-188509/23.
DR N-PSDB: Q48393.
PT Recombinant human gene encoding human pseudo-cholesterase -
PT used to treat organo-phosphorus poisoning
PS Claim 4; Columns 41-46; 34pp; English.
CC A cDNA library prepared from foetal brain mRNA was screened with
CC degenerate probe pools based on the organophosphorus binding site of
CC cholinesterases. A 764 nucleotide insert (designated FBChE12) was
CC isolated from one positive clone and sequenced. This insert (Q42495),
CC containing an ORF large enough to code for about half the subunit
CC size of human cholinesterase, was used as a probe to obtain the full-
CC length pseudocholesterase sequence by screening a foetal liver
CC cDNA library and a primary glioblastoma cDNA library (Q48393).
CC This sequence differs from sequence Q42496, particularly in the
CC 3'-UTR.
SQ Sequence 635 AA;


Alignment_scores:  

  Quality: 60.50 Length: 62  

  Ratio: 1.779 Gaps: 5  

  Percent Similarity: 54.839 Percent Identity: 29.032



Alignment_block:  

  US-09-030-606-115 x R41509 ..



Align seg 1/1 to: R41509 from: 1 to: 635



```

      8 CTCCTCCCTCCTGATTAATTCATTGCAATTCGCAAGGANT 57
      ::::::::::: ::::::::::: :::::::::::
     10 IIESEPRDASNTYTYTMeIlePheHnProCysLysVal..... 23
     58 ACACATTCTACTGTGATGTATGTGTGCAAAAAAAGTGTCT 107
     24 .....CysHISLeuCyScysArgGluSerGluIleAsnM 35
     108 TGTGT...AAATTACT.....TGGTTTGTG 130
      ::::::::::: ::::::::::: :::::::::::
     35 eCHISerLysValThrIleIleCysIleArgPheLeuPheTrpPheVal 51
     131 AATTCATCTTGCTTTTCCCATTCGCACTAGTCAT 166
     52 ...LeuLeuCyScysMetLeu...IleGlyLysSerHis 61

```



seq_name: A_Geneseq_36:R24297



seq_documentation_block:  

  ID R24297 standard; Protein; 322 AA.  

  AC R24297;  

  DT 20-NOV-1992 (first entry)  

  DE Glycopeptide resistance protein VanH from E.faecium.  

  KW Glycopeptide antibiotic; vancomycin; telcoplanin; resistant;  

  KW D-Ala-D-Ala ligase; peptidoglycan precursor; transposon;  

  KW inverted repeats.  

  OS Enterococcus faecium BM4147.  

  PN W09207942-A.  

  PD 14-MAY-1992.  

  PF 29-OCT-1991; F00855.


```

PK 07-MAR-1989; EP-104023.  
PB 24-MAY-1989; EP-109656

FT /note="u

Line	Description	Amount	Balance
1	FT	1.00	1.00
2	Misc_difference	287	288

FT /note= "unspecified amino acid"  
 PN WC0854206-A1.  
 PD 03-DEC-1998.  
 PF 28-MAY-1998: U10868.  
 PR 28-AUG-1997: US-056296.  
 PR 30-MAY-1997: US-044039.  
 PR 30-MAY-1997: US-048093.  
 PR 30-MAY-1997: US-048101.  
 PR 30-MAY-1997: US-048190.  
 PR 30-MAY-1997: US-048356.  
 PR 30-MAY-1997: US-050935.  
 PR 29-AUG-1997: US-056250.  
 PR 29-AUG-1997: US-056293.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Carter KC, Dillon PJ, Endress GA, Feng P, Ni J,  
 PI Rosen CA, Ruben SM, Yu G;  
 DR WPI: 99-070209/06.  
 DR N-PSDB: V08819.  
 PT New isolated human genes - useful for diagnosis and treatment of,  
 PT e.g. cancers, neurological disorders, immune diseases, developmental  
 PT disorders or blood disorders  
 PS Claim 11, Page 151-152, 188pp: English.  
 CC This sequence is encoded by a cDNA of the invention, designated  
 CC Gene No. 9. This sequence represents a human secreted protein, and is  
 CC expressed in foetal liver, spleen and pituitary gland and to a  
 CC lesser extent in multiple tissues.  
 CC The DNA sequences of the invention and their corresponding secreted  
 CC polypeptides are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the DNA sequences.  
 CC Specific uses are described for each of the DNA sequences and the encoded  
 CC proteins, based on which tissues they are most highly expressed in, and  
 CC include developing products for the diagnosis or treatment of cancer,  
 CC tumours, neurological disorders, developmental abnormalities and foetal  
 CC deficiencies, blood disorders, leukemias, diseases of the immune system  
 CC (including allergies or asthma), hepatic disease, Alzheimer's and  
 CC cognitive disorders, schizophrenia, prostate diseases, autoimmune  
 CC disorders and AIDS. The polypeptides are also useful for identifying  
 CC their binding partners.  
 SQ Sequence 287 AA;

Alignment\_scores:  
 Quality: 59.00 Length: 109  
 Ratio: 1.204 Gaps: 4  
 Percent Similarity: 44.954 Percent Identity: 21.101

Alignment\_block:  
 US-09-030-606-115 x W73405 ..

Align seg 1/1 to: W73405 from: 1 to: 287

```

20 CTGAATTATCTTCACTGCAATTTGCAAGATTACATTTCAC 69
||| :||| :||| :||| :||| :||| :||| :||| :|||
6 LeuYrMetValMetSerAlaCysPheLeu...AlaAlaGlyIlePheYr 21
70 GTGATGTATATTGTTGTCGCAAAAAAAGTCTTGTGTTAAATTA 119
||| :||| :||| :||| :||| :||| :||| :||| :|||
21 PValSerIleLeuYs....ArgAsnThrTyrSerValPheLysIleH 36
120 CTGTGGTTGTAATCCATCTTCTTTCCCATGGAAGTATGATTTAA 169
||| :||| :||| :||| :||| :||| :||| :||| :|||
36 IStrPLeuMetAlaAlaLeuAlaPhe..... 44
170 CCCATCTCTGAACTGTAGAAAAACATCTGAAGCTAGTCTATCAGCAT 219
||| :||| :||| :||| :||| :||| :||| :||| :|||
45 .....ThrLysSerIleSerLeuLeuPhe 52
220 CTGACAGGTGAATGTGATGTTCTCAGAACATTTCCACCCAGACAGCTG 269
||| :||| :||| :||| :||| :||| :||| :||| :|||
52 eHtIsSerIleAsnTyrTyrPheIleAsnSerGlnGlyPro..... 65

```

270 TTTCATCTGTGTTAATAAATTAGTTGGTTCCTCATCATCAACA 319  
 ||| :||| :||| :||| :||| :||| :||| :||| :|||  
 66 .....ProHisArgArg 69  
 320 CCCCTGCTCCAAATCTGTCAATPAAAAGT 346  
 ||| :||| :||| :||| :||| :||| :||| :||| :|||  
 70 ProCysArgHisValLeuHisArgPhe 78

seq\_name: A\_Geneseq\_36:R24307

seq\_documentation\_block:

ID R24307 standard; Protein; 2408 AA.

AC R24307;

DE Translation of ORF 3 contig. E.faecium proteins VanH and VanX.

KW Glycopeptide antibiotic; Vancomycin; Telicoplanin; resistant;

KW D-Ala-D-Ala ligase; peptidoglycan precursor; transposon;

KW Inverted repeats; vanR; vanS; vanH; vanA; vanX;

KW open reading frame.

OS Enterococcus faecium BM4147.

FH key Location/Qualifiers

FT protein 1167..1510

FT /label= VanH

FT /note= "see R24297"

FT /label= VanX

FT /note= "see R24299"

PN WC09207942-A.

PD 14-MAY-1992.

PF 29-OCT-1991: F00855.

PR 31-OCT-1990: FR-013579.

PA (INSP ) INST PASTEREUR.

PI Arthur M, Courvalin P, Dutka-malen S, Molinas C;

DR WPI: 92-183677/22.

DR N-PSDB: Q25183.

PT Polypeptides involved in expression of glycopeptide antibiotic

PT resistance - useful in diagnosing presence of Gram-positive

PT enterococcal strains e.g. Enterococcus faecium and E.gallinarum

PS Disclosure; Fig 5; 163pp; French.

CC A 7.3kb fragment of E.faecium DNA containing the five genes vanH,

CC vanA, vanX, vanR and vanS involved in antibiotic resistance was

CC translated in each of the three possible open reading frames.

CC Within ORF 3 were the vanH and vanX genes. The "X"s in the sequence

CC indicate the position of nonsense codons.

CC See also Q25178-Q25182.

SQ Sequence 2408 AA;

Alignment\_scores:  
 Quality: 60.00 Length: 88  
 Ratio: 1.250 Gaps: 4  
 Percent Similarity: 54.545 Percent Identity: 25.000

Alignment\_block:  
 US-09-030-606-115/rev x R24307 ..

Align seg 1/1 to: R24307 from: 1 to: 2408

```

276 ATAGAACAGAGCTGTGCTGGAATGTTGAGAACATCCAA...TT 230
::| :||| :||| :||| :||| :||| :||| :||| :|||
1339 ValGlyThrGlyGlnIleGlyLysAlaValIleGluArgLeuArgGlyPhe 1355
229 CACCTGTCAAGATGCTGATAGACTAGCTCTTCAGATGTTTCTACCACTT 180
||| :||| :||| :||| :||| :||| :||| :||| :|||
1355 eGlyCysLysValLeu.....AlaTyrSerArgSerA 1366
179 CAGAGATGGTTAATGACTAGTCCATGGGG.....AAAAGCAA 139
||| :||| :||| :||| :||| :||| :||| :||| :|||
1366 rGserIleGluValAsnTyrValProPheAspGlyLeuLeuGlnAsnSer 1382
138 GATGATTACAAACCAAGTAAATTTAAACAAGACACTTTTCTTTT 89
||| :||| :||| :||| :||| :||| :||| :||| :|||
1383 AspIleValThrLeuHisValProLeuAsnThrAspThrHisTyrIleIle 1399

```



```

      679  :|||  |||  |||::  :::::|||  |||  |||
      679  tleuvalprolaltprgin...Glnvalthrproleualproalaltprt 695
185  GTAAAGAAACATCTGAGAGCTAGCTCTACGATCTGACAGCGAATGT 234
      695  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
695  hrthrleuthrsercluservalaaglyserhlsargleuglyasp.tr 711
235  GATGCTCTC..... 244
      711  :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
711  polylysmethlesecysserasnhtyrasnservalmetproginp 728
245  .....AGAACATTTCACCCAGACACCTGTTTTCATC 277
      728  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
728  rolauleuthrasnglnlthrleuSerAlaProGlnProValSerVal 744
278  CTGTTTAAATAGTTTG.....GTCTCTACATGATACAA 318
      745  :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
745  glyllealaahlsvalvaltrproginproalathrthlylsasny 761
319  ACCCTGCTCAATCTGCACAT 340
      761  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
761  sclncysclnasmargserasn 768

```

seq\_name: A\_Geneseq\_36:p60007

seq\_documentation\_block:

ID P60007 standard: Protein; 278 AA.  
 DT 30-JUL-1991 (first entry)  
 DE Sequence of oncogene v-ras.  
 KW Therapy; diagnosis; insulin.  
 PN EP-192392-A.  
 PD 27-AUG-1986.  
 PR 10-FEB-1986; 300894.  
 PR 12-FEB-1985; US-700776.  
 PA (GETH ) GENENTECH INC.  
 PI Bell JR, Ulrich A, Ramchandran J;  
 DR WPI: 86-226966/35.  
 PT New DNA encoding insulin receptor or its fragments - used for  
 synthesis of receptor and mutants for therapeutic and diagnostic  
 use  
 PS Example: Fig 5; 62pp; English.  
 CC A mutant IR is claimed which may have a mutated alpha-chain, esp. at  
 CC the precursor processing site. The beta-chain may be mutated, e.g.  
 CC by deletion of the transmembrane sequence; the tyrosine kinase  
 CC activity may be inactivated. Fig. 5 is a comparison of oncogene and  
 CC human EGF receptor sequences with that of HIR in the cytoplasmic  
 CC domain of the insulin receptor beta subunit.  
 SQ Sequence 278 AA;

alignment\_scores:

Quality:	56.50	Length:	21
Ratio:	3.531	Gaps:	1
Percent Similarity:	76.190	Percent Identity:	47.619

alignment\_block:

US-09-030-606-115/rev x p60007 ..

Align seg 1/1 to: P60007 from: 1 to: 278

```

111 AACAAAGACTTTTGTTCACAC.....AATATACATCA 71
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
255 AsnArgProThrPhePheTyrIleHisLysLeuGlnIleArgHl 271
70  CAGTCAATGTCT 58
      ||||  ::|||
271 sSerProleucys 275
seq_name: A_Geneseq_36:w97634
seq_documentation_block:
ID W97634 standard: Protein; 281 AA.

```

```

AC W97634;
DT 10-MAY-1999 (first entry)
DE Xenopus Frazzled protein W67-16.
KW Frazzled; W67-16; cell growth; cell differentiation;
KW cell proliferation; wound healing; vulnery; tissue repair;
KW pancreatitis; diabetes; rheumatoid arthritis; osteoarthritis;
KW chondrocyte; cartilage; therapy.
OS Xenopus sp.
PN MO9902679-A1.
PD 21-JAN-1999.
PE 28-APR-1998; U08517.
PR 11-JUL-1997; US-893654.
PA (GENE ) GENETICS INST INC.
PA (WHEH ) WHITEHEAD INST BIOMEDICAL RES
PI Lavallie ER, Paulsen J, Racie L, Sive H, Sun B;
DR WPI: 99-120880/10.
DR N-PSDB; X06998.
PT New DNA encoding Frazzled proteins - for modulating growth,
PT differentiation, proliferation of cells, e.g. in treatment of
PT wounds, diabetes and rheumatoid arthritis
PS Claim 3; Page 44-45; 71pp; English.
CC This is the amino acid sequence of novel Xenopus frazzled protein
CC W67-16, as deduced from an isolated cDNA clone (see X06998). The
CC invention provides Xenopus W67-16, W67-19 and W628-5 polypeptides
CC (see W98634-36), isolated DNA molecules (see X06998-700), vectors,
CC host cells and methods for recombinant protein production.
CC Frazzled proteins are used: to modulate regulation of pancreatic
CC genes (e.g. for pancreatic tissue repair and for treatment or
CC prevention of pancreatic tumours, diabetes, pancreatitis etc.,
CC including inducing de novo formation of beta-islet cells); to
CC modulate cell development (particularly in haematopoietic, gut,
CC neuronal or muscle tissue, including expression from the nucleic
CC acid in vivo), particularly for inducing cell formation, growth,
CC differentiation, proliferation and maintenance, specifically
CC induction of chondrocytes and/or cartilage tissue by treating
CC progenitor cells in vitro; generally to modulate effect of wnt
CC genes; to raise antibodies; and as dietary supplement or cell
CC culture additive. They are also used: to treat diseases associated
CC with defects in cell growth, formation etc., also for augmenting
CC the activity of other factors with regenerative/differentiation
CC activities, e.g. in wound healing; to treat or prevent rheumatoid
CC arthritis, osteoarthritis or other articular cartilage disorders;
CC in combination with transplants of cells for treating neuronal
CC death; and to induce formation of hormone-producing cells, in both
CC human and veterinary medicine.
SQ Sequence 281 AA;

```

alignment\_scores:

Quality:	56.50	Length:	31
Ratio:	2.690	Gaps:	1
Percent Similarity:	67.742	Percent Identity:	32.258

alignment\_block:

US-09-030-606-115/rev x W97634 ..

Align seg 1/1 to: W97634 from: 1 to: 281

```

123 CAACTAATTTTAAACAAGACACTTTTGTTCACACAAATATACA 74
      :|||  |||  |||  |||  |||  |||  |||  |||  |||  |||
234 LysleuileargasnargprothvaltyrvallelaaglyaspIleHl 250
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
73 TCACAGTGAATG.....TGTAACTCTGCAAATTCGAA 40
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
250 sHsGlyLysIleLysIlePheCysSerProCysValleuArg 264
seq_name: A_Geneseq_36:w06503
seq_documentation_block:
ID W06503 standard: Protein; 214 AA.
AC W06503;
DT 06-FEB-1997 (first entry)
DE Htm4 protein.

```





This Page Blank (uspto)





Alignment\_block:  
US-09-030-606-115/rev x US-08-286-819A-2 ..  
Align seg 1/1 to: US-08-286-819A-2 from: 1 to: 322

```

276 ATGAAACAGCGCTGCTGGTGAATGTTCTGAGAACATCCAA...TT 220
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
152 ValGlyThrGlyGlnIleGlyLysAlaValIleGluArgLeuArgGlyPn 168
229 CACCTGTACAGTGTATGATGACTAGCTCTTCAGATGTTTCTACAGGT 180
    |::|::|::|::|::|::|::|::|::|::|::|::|::|
168 eGlyCysLysValLeu.....AlaTyrSerArgSerA 179
179 CAGAGATGGGTAAATGACTAGTTCATGCGG.....AAAAAGCA 139
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
179 rGserIleGluValAsnTyrValProPheAspLueLeuGlnAsnSer 195
138 GATGATTCACAACCAAGTAATTTAAACAAGACACTTTTCTTTT 89
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|
196 AspIleValThrLeuHisValProLeuAsnThrAspThrHisTyrIleI 212
88 GCACACAAATATACATCACAGTGAATGTCTATCCTTGCAATGCAAG 39
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
212 eSerHisGlu.....GlnI 217
38 TTGAAGAATTAA 25
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
217 leGlnArgMetLys 221

```

seq\_name: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:US-08-286-819A-29

seq\_documentation\_block:  
Sequence 29, Application US/08286819A  
Patent No. 5871910

```

GENERAL INFORMATION:
APPLICANT: ARTHUR, MICHEL
APPLICANT: DUKTA-MALEN, SYLVIE
APPLICANT: MOLINAS, CATHERINE
APPLICANT: COURVALIN, PATRICE
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS, IN PARTICULAR
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESSES:
ADDRESS: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,
STREET: 1755 S. Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,819A
FILING DATE: 05-AUG-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/174,682
FILING DATE: 28-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/917,146
FILING DATE: 10-AUG-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR/91/00855
FILING DATE: 29-OCT-1991
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: FR 9013579
FILING DATE: 31-OCT-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NO. 5871910man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 660-060-0 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 ORAT UR
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 2291 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-286-819A-29

```

alignment\_scores:  
Quality: 60.00 Length: 88  
Ratio: 1.250 Gaps: 4  
Percent Similarity: 34.545 Percent Identity: 25.000

alignment\_block:

US-09-030-606-115/rev x US-08-286-819A-29 ..

Align seg 1/1 to: US-08-286-819A-29 from: 1 to: 2291

```

276 ATGAAACAGCGCTGCTGGTGAATGTTCTGAGAACATCCAA...TT 230
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1264 ValGlyThrGlyGlnIleGlyLysAlaValIleGluArgLeuArgGlyPn 1280
229 CACCTGTACAGTGTATGATGACTAGCTCTTCAGATGTTTCTACAGTT 180
    |::|::|::|::|::|::|::|::|::|::|::|::|::|
1280 eGlyCysLysValLeu.....AlaTyrSerArgSerA 1291
179 CAGAGATGGGTAAATGACTAGTTCATGCGG.....AAAAAGCA 139
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1291 rGserIleGluValAsnTyrValProPheAspLueLeuGlnAsnSer 1307
138 GATGATTCACAACCAAGTAATTTAAACAAGACACTTTTCTTTT 89
    |||::|::|::|::|::|::|::|::|::|::|::|::|::|
1308 AspIleValThrLeuHisValProLeuAsnThrAspThrHisTyrIleI 1324
88 GCACACAAATATACATCACAGTGAATGTCTATCCTTGCAATGCAAG 39
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1324 eSerHisGlu.....GlnI 1329
38 TTGAAGAATTAA 25
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1329 leGlnArgMetLys 1333

```

seq\_name: /cgn2\_6/ptodata/2/1aa/PCTUS9\_COMB.pep:PCT-US94-14436-28

seq\_documentation\_block:

Sequence 28, Application PC/TUS9414436

GENERAL INFORMATION:

APPLICANT: Baird, Jonathan A.

APPLICANT: Walker, Mary

APPLICANT: Branchek, Theresa

APPLICANT: Weinschenk, Richard L.

TITLE OF INVENTION: DNA ENCODING A HUMAN NEUROPEPTIDE

TITLE OF INVENTION: Y/PEPTIDE YY/PANCREATIC POLYPEPTIDE

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

```

: ZIP: 10036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US94/14436
: FILING DATE:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: White, John P.
: REGISTRATION NUMBER: 28,678
: REFERENCE/DOCKET NUMBER: 44743-A-PCT\JPM\MAT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 278-0400
: TELEFAX: (212) 391-0525
: INFORMATION FOR SEQ ID NO: 28:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 375 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US94-14436-28

```

```

Alignment_scores:
Quality: 58.00 Length: 84
Ratio: 1.349 Gaps: 5
Percent Similarity: 51.190 Percent Identity: 29.762

```

```

Alignment_block:
US-09-030-606-115 x PCT-US94-14436-28 ..

```

```

Align seg 1/1 to: PCT-US94-14436-28 from: 1 to: 375

```

```

33 TTTCACCTTGCAATTTGCAAGATTACACATTTCACTGTGATGATATG 82
||| :|||:||||| :|||: ||| :|||
109 PheGylgluValleucysylsMetleuthrpe.....IleGlnCy 122
83 TGTTCACAAAAAAGAGTCTTGTGTTAAATTAATTACTGTGTTGTGA 132
|||:||||| :||| :||| :|||:|||||
122 SmetSerValThrValSerIleLeuSerleuValleuValAlaIleGluA 139
133 TCCATCTTCCTTTTCCCATGGAACATGATCATTAACCATCTCGAAC 182
||| :|||:||||| :|||: ||| :|||
139 rghIs.....GlnLeuIleIleasnProthr...Gly 148
183 TGGTAGAAAAACATCTGAAGACTAGTCTATCAGCATCTGACAGTGAA 232
||| :|||:||||| :|||: ||| :|||
149 TrpLysProSerIleSerGlnAlaTrpLeuGlyIleVal.....Va 162
233 TGGATGCTTC.....TCAGACCATTTCCACCGACAGCC 267
: ||||| :||| :|||:||||| :|||:
162 IleTrpPheIleSerCysPheLeuSerleuProPheLeuAlaAsnSerI 179
268 TG 269
179 le 179

```

```

seq_name: /cgn2_6/ptodata/2/1ae/5A_COMB.pep:US-08-318-492-4

```

```

seq_documentation_block:
: Sequence 4, Application US/08318492
: Patent No. 5552312
: GENERAL INFORMATION:
: APPLICANT: Lim, Bing
: APPLICANT: Adira, Chaker N.
: APPLICANT: Lellias, Jean-Michel
: TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND
: TITLE OF INVENTION: ASSAYS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02173

```

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/318,492
: FILING DATE:
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Brook, David E.
: REGISTRATION NUMBER: 22,592
: REFERENCE/DOCKET NUMBER: BIH94-03
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 861-9540
: TELEFAX: (617) 861-9540
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 214 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-318-492-4

```

```

Alignment_scores:
Quality: 56.00 Length: 62
Ratio: 1.536 Gaps: 4
Percent Similarity: 58.065 Percent Identity: 37.097

```

```

Alignment_block:
US-09-030-606-115/rev x US-08-318-492-4 ..

```

```

Align seg 1/1 to: US-08-318-492-4 from: 1 to: 214

```

```

253 AATGCTTCAGAACCAATTCACCTGTCA...GATGCTGATGACT 207
||| ||| |||||:||||| :||| ||| :|||:
30 AsnThrSerValThrHisProIleasnGlySerProAspTyrGlnLysAl 46
206 AGCTTTCAGATGTT...TCTACCACTTCAGAGATGGGTTAATGACTA 160
|||:||||| :|||: ||| :|||
46 alysLeuGlnValleuGlyAlaIleGlnIleLeuAsnAlaIleMetIleL 63
159 GTTCCAAATGGGAAAAAGCAGATGATTCACAAACCAAGTAATTAA 110
: ||| :||| :|||:||||| :|||:
63 euAlaLeuGlyValPheLeuGlySerLeuGlnTyrProTyrHisPhe... 78
109 CAAGACACTTTTGTGCAACACAAATATACA 74
|||:||||| :|||: ||| :|||
79 GlnLysHisPhePhePhe...ThrPheTyrThr 89

```

```

seq_name: /cgn2_6/ptodata/2/1ae/5A_COMB.pep:US-08-707-340-4

```

```

seq_documentation_block:
: Sequence 4, Application US/08707340
: Patent No. 5705615
: GENERAL INFORMATION:
: APPLICANT: Lim, Bing
: APPLICANT: Adira, Chaker N.
: APPLICANT: Lellias, Jean-Michel
: TITLE OF INVENTION: RECOMBINANT HTM4 GENE, PROTEIN AND
: TITLE OF INVENTION: ASSAYS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington

```



```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: 2 Millie Drive
CITY: Lexington
STATE: MA
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/03076
FILING DATE: 19930331
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WHI92-03A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1513 amino acids
TYPE: AMINO ACID
TOPOLOGY: Linear
MOLECULE TYPE: protein
PCT-US93-03076-2

alignment_scores:
Quality: 57.00 Length: 92
Ratio: 1.056 Gaps: 3
Percent Similarity: 58.696 Percent Identity: 22.826

alignment_block:
US-09-030-606-115 x PCT-US93-03076-2 ..
Align seg 1/1 to: PCT-US93-03076-2 from: 1 to: 1513
19 TCTGATTTAATCTTCACTTGCATTTGCAAGATTACACATTTCAC 68
||| :|||:||||:||||:||||:||||:||||:||||:||||:
839 SerTyrHisSerSerPheSerIleArgLysSerArgLeuValHisGly 855
69 TGTGATGATATTGTGTGCAAAAAAAGTCTTTGTTTAAATT 118
:||||:||||:||||:||||:||||:||||:||||:||||:
835 rIleValPheTyrSerAlaLysArgLysAlaSerLeuAlaMetLeuArg 872
119 ACTTGCTTTGATTCATCTGCTTTTCCCATTTGGAACATGATCATTA 168
||| :|||:||||:||||:||||:||||:||||:||||:
872 lAphLeuLysGluValGlnAspIleIleProIleGlnLeuValAlaLeu 888
169 ACC.....CATCTCTGAACCTGCTAGAAAAACATCTGAAGAGCT 206
||| :|||:||||:||||:||||:||||:||||:||||:
889 ThrAspGlyAlaIleAspValLeuAspAsnAspLeuSerArgGluGln 905
207 ACTCTAT..CAGCATCTGACAGGTGAATGGATGGTCTCAGAACCAT 253
||| :|||:||||:||||:||||:||||:||||:||||:
905 uThrGluGlyGluGluIleAlaGlnGluIleAspGly.....ArgP 919
254 TCACCCAGACAGCCTGTTCTATCCT 279
|||:||||:||||:||||:||||:||||:||||:
919 MetHisSerIleProCysSerHisPro 927

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-564-458-1
seq_documentation_block:
; Sequence 1, Application US/08564458
; Patent No. 5641749
; GENERAL INFORMATION:
; APPLICANT: Yan, Qiao
```

```

TITLE OF INVENTION: Method for Treating Retinal Ganglion Cell Injury
TITLE OF INVENTION: Using Gial Cell Line-Derived Neurotrophic
TITLE OF INVENTION: Factor (GDNF) Protein Product
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: United States of America
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/564,458
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Curry, Daniel R.
REGISTRATION NUMBER: 32,727
REFERENCE/DOCKET NUMBER: A-360
TELECOMMUNICATION INFORMATION:
TELEPHONE: 805-447-8102
TELEFAX: 805-499-8011
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 134 amino acid residues
TYPE: amino acid
TOPOLOGY: Linear
FEATURE:
NAME/KEY: Inferred amino acid sequence for mature human GDNF
US-08-564-458-1

alignment_scores:
Quality: 54.00 Length: 40
Ratio: 2.000 Gaps: 1
Percent Similarity: 67.500 Percent Identity: 30.000

alignment_block:
US-09-030-606-115 x US-08-564-458-1 ..
Align seg 1/1 to: US-08-564-458-1 from: 1 to: 134
136 ACTTGCTTTTCCCATTTGGAAGTCAATTAACCATCTGTGAACCTG 185
||| :|||:||||:||||:||||:||||:||||:||||:
42 VALLeuThrAlaIleHisLeuAsn.....ValThrAspLeuGlyLeuGI 56
186 TAGAAAAACATCTGAAGAGTAGTATACATCATCTGACAGTGAATTGG 235
| :|||:||||:||||:||||:||||:||||:||||:
56 YTyrcIuThrLysGluGluIleuIlePheArgTyrCysSerGlySerCys 73
236 ATGCTTTCAGAACCATTTTC 255
|||:||||:||||:||||:||||:||||:||||:
73 spAlaAlaGluThrThrTyr 79

seq_name: /cgn2_6/prodata/2/1aa/5A_COMB.pep:US-08-564-833-1
seq_documentation_block:
; Sequence 1, Application US/08564833
; Patent No. 5641750
; GENERAL INFORMATION:
; APPLICANT: Louis, Jean-Claude
TITLE OF INVENTION: Methods for Treating Photoreceptors using Gial Cell
TITLE OF INVENTION: Line-Derived Neurotrophic Factor (GDNF) Protein
PRODUCT
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: AMGEN INC.
```

```

: STREET: 1840 DeHavilland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: United States of America
: ZIP: 91320
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/564,833
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Curry, Daniel R.
: REGISTRATION NUMBER: 32,727
: REFERENCE/DOCKET NUMBER: A-363
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 805-447-8102
: TELEFAX: 805-499-8011
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 134 amino acid residues
: TYPE: amino acid
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: Inferred amino acid sequence for mature human GDNF
: US-08-564-833-1

```

```

alignment_scores:
  Quality: 54.00      Length: 40
  Ratio: 2.000      Gaps: 1
  Percent Similarity: 67.500      Percent Identity: 30.000

```

```

alignment_block:
  US-09-030-606-115 x US-08-564-833-1  ..

```

```

Align seg 1/1 to: US-08-564-833-1 from: 1 to: 134

```

```

136 ATCTGCTTTTCCCATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGG 185
   |||||      |||      |||      |||      |||      |||
42 ValLeuThrAlaIleHisLeuAsn.....ValThrAspLeuGlyLeuG1 56
186 TAGAAAAACATCTGAAGAGCTAGTCTATCAGACATCTGACAGTGAATTGG 235
   |  ::|||::|||::|||::|||::|||::|||::|||::|||::|||::
56 YTYGluThrLysGluGluLeuIlePheArgTYrCysSerGlySerCysA 73

```

```

236 ATGCTTCTCAGAACCATTTTC 255
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
73 spAlaAlaGluThrThyTyr 79

```

```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-535-682-1

```

```

seq_documentation_block:

```

```

: Sequence 1, Application US/08535682
: Patent No. 5731284
: GENERAL INFORMATION:
: APPLICANT: Williams, Lawrence R.
: TITLE OF INVENTION: Method for Treating Alzheimer's Disease
: TITLE OF INVENTION: Using Gial Cell Line-Derived Neurotrophic
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: AMGEN INC.
: STREET: 1840 DeHavilland Drive
: CITY: Thousand Oaks
: STATE: California
: COUNTRY: United States of America
: ZIP: 91320
: COMPUTER_READABLE FORM:

```

```

: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/535,682
: FILING DATE:
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Curry, Daniel R.
: REGISTRATION NUMBER: 32,727
: REFERENCE/DOCKET NUMBER: A-358
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 805-447-8102
: TELEFAX: 805-499-8011
: TELEX:
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 134 amino acid residues
: TYPE: amino acid
: TOPOLOGY: linear
: FEATURE:
: NAME/KEY: Inferred amino acid sequence for mature human GDNF
: US-08-535-682-1

```

```

alignment_scores:
  Quality: 54.00      Length: 40
  Ratio: 2.000      Gaps: 1
  Percent Similarity: 67.500      Percent Identity: 30.000

```

```

alignment_block:
  US-09-030-606-115 x US-08-535-682-1  ..

```

```

Align seg 1/1 to: US-08-535-682-1 from: 1 to: 134

```

```

136 ATCTGCTTTTCCCATGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGG 185
   |||||      |||      |||      |||      |||      |||
42 ValLeuThrAlaIleHisLeuAsn.....ValThrAspLeuGlyLeuG1 56
186 TAGAAAAACATCTGAAGAGCTAGTCTATCAGACATCTGACAGTGAATTGG 235
   |  ::|||::|||::|||::|||::|||::|||::|||::|||::|||::
56 YTYGluThrLysGluGluLeuIlePheArgTYrCysSerGlySerCysA 73
236 ATGCTTCTCAGAACCATTTTC 255
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
73 spAlaAlaGluThrThyTyr 79

```

```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-446-383A-1

```

```

seq_documentation_block:

```

```

: Sequence 1, Application US/08446383A
: Patent No. 5733875
: GENERAL INFORMATION:
: APPLICANT: Martin, David
: TITLE OF INVENTION: METHODS OF USING GDNF AS A
: TITLE OF INVENTION: NEUROPROTECTIVE AGENT
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: AMGEN INC.
: STREET: 1840 DeHavilland Drive
: CITY: Thousand Oaks
: STATE: CA
: COUNTRY: US
: ZIP: 91320-1789
: COMPUTER_READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/446,383A
: FILING DATE: 22-MAY-1995

```



TELEFAX: (314) 727-6092  
INFORMATION FOR SEQ ID NO: 76:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 134 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-519-777-76

alignment\_scores:  
Quality: 54.00 Length: 40  
Ratio: 2.000 Gaps: 1  
Percent Similarity: 67.500 Percent Identity: 30.000

alignment\_block:  
US-09-030-606-115 x US-08-519-777-76 ..

Align seg 1/1 to: US-08-519-777-76 from: 1 to: 134

136 ATCTGCTTTTCCCATGAGACTAGTCATTAACCATCTGACTGG 185  
:::||||| ||| ||| :::||| ||| |||  
42 ValLeuThrAlaIleHisLeuAsn.....ValThrAspLeuGlyLeuG1 56  
186 TAGAAAAACATCTGAAGAGCTAGTCTATCAGCATCTGACAGGTGAATTGG 235  
| :::::::::::::::::::::  
56 yTyrgIuThrLySGluGluLeuIlePheArgTyrcysSerGlySerCysa 73  
236 ATGCTTCTCAGAACCATTTC 255  
| :::::::::::::::::::::  
73 spAlaIaGluThrThrTyr 79

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-618-543-1

seq\_documentation\_block:

Sequence 1, Application US/08618543  
Patent No. 5741778  
GENERAL INFORMATION:  
APPLICANT: Martin, David  
APPLICANT: Miller, Gerald D.  
TITLE OF INVENTION: Method for Treating Huntington's Disease  
TITLE OF INVENTION: Using Glial Cell Line-Derived Neurotrophic Factor (GDNF)  
TITLE OF INVENTION: Protein Product  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: AMGEN INC.  
STREET: 1840 DeHarvilland Drive  
CITY: Thousand Oaks  
STATE: California  
COUNTRY: USA  
ZIP: 91320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/618,543  
FILING DATE:  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Curry, Daniel R.  
REFERENCE/DOCKET NUMBER: A-390  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 134 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-618-543-1

alignment\_scores:  
Quality: 54.00 Length: 40  
Ratio: 2.000 Gaps: 1  
Percent Similarity: 67.500 Percent Identity: 30.000

alignment\_block:  
US-09-030-606-115 x US-08-618-543-1 ..

Align seg 1/1 to: US-08-618-543-1 from: 1 to: 134

136 ATCTGCTTTTCCCATGAGACTAGTCATTAACCATCTGACTGG 185  
:::||||| ||| ||| :::||| ||| |||  
42 ValLeuThrAlaIleHisLeuAsn.....ValThrAspLeuGlyLeuG1 56  
186 TAGAAAAACATCTGAAGAGCTAGTCTATCAGCATCTGACAGGTGAATTGG 235  
| :::::::::::::::::::::  
56 yTyrgIuThrLySGluGluLeuIlePheArgTyrcysSerGlySerCysa 73  
236 ATGCTTCTCAGAACCATTTC 255  
| :::::::::::::::::::::  
73 spAlaIaGluThrThrTyr 79

seq\_name: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:US-08-742-035-76

seq\_documentation\_block:

Sequence 76, Application US/08742035  
Patent No. 5747655  
GENERAL INFORMATION:  
APPLICANT: JOHNSON JR., EUGENE M.  
APPLICANT: MILBRANDT, JEFFREY D.  
APPLICANT: KOTZBAUER, PAUL T.  
APPLICANT: LAMPE, PATRICIA A.  
TITLE OF INVENTION: NEURUFURIN AND RELATED GROWTH FACTORS  
NUMBER OF SEQUENCES: 78  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ROGERS, HOWELL & HAFERKAMP, L.C.  
STREET: 7733 FORSYTH BOULEVARD, SUITE 1400  
CITY: ST. LOUIS  
STATE: MISSOURI  
COUNTRY: US  
ZIP: 63105-1817  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/742,035  
FILING DATE: 01-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/519,777  
FILING DATE: 28-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: HOLLAND, DONALD R.  
REGISTRATION NUMBER: 35,197  
REFERENCE/DOCKET NUMBER: 953095  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (314) 727-5188  
TELEFAX: (314) 727-6092  
INFORMATION FOR SEQ ID NO: 76:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 134 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-742-035-76

alignment\_scores:  
Quality: 54.00 Length: 40  
Ratio: 2.000 Gaps: 1  
Percent Similarity: 67.500 Percent Identity: 30.000



alignment\_block:  
US-09-030-606-115 x US-08-742-035-76 ..

Align seg 1/1 to: US-08-742-035-76 from: 1 to: 134

```
136 ATCTTGCTTTTCCCATTTGGAAGTAGTCATTACCCATCTCTGAAGTGG 185
   ::|||      ||      ||      ||      ||      ||
42 ValLeuThrAlaIleHisLeuasn.....ValThrAspLeuGlyLeuGI 56
   *
186 TAGAAAACATCTGAAGAGCTAGCTATCAGCATCTGACAGGTGAATTGG 235
   |  ::|||::|||::|||::|||::|||::|||::|||::|||::|||
56 yTyrGIuThrLysGIuGIuLeuIlePheArgTyrCysSerGIySerCysA 73
   |  ::|||::|||::|||::|||::|||::|||::|||::|||
236 ATGGTTCAGAACCATTTTC 255
   ||::|||::|||::|||::|||::|||::|||::|||
73 sPAlaIaGIuThrThrTyr 79
```

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:21:38 ; Search time 2928.69 seconds  
(without alignments)  
397.445 Million cell updates/sec

Title: US-09-030-606-115  
Perfect score: 366  
Sequence: 1 GCTCTTCTCTCCCTCCTC.....CTGACTGAGATTAGTC 366

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_ba1: \*  
2: gb\_ba2: \*  
3: gb\_com: \*  
4: gb\_ov: \*  
5: gb\_pat: \*  
6: gb\_ph: \*  
7: gb\_pl1: \*  
8: gb\_pl2: \*  
9: gb\_pl3: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_v1: \*  
18: em\_fun: \*  
19: em\_htg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_com: \*  
24: em\_ov: \*  
25: em\_pat: \*  
26: em\_ph: \*  
27: em\_pl: \*  
28: em\_ro: \*  
29: em\_sts: \*  
30: em\_sy: \*  
31: em\_un: \*  
32: em\_v1: \*  
33: gb\_htg1: \*  
34: gb\_htg2: \*  
35: gb\_in1: \*  
36: gb\_in2: \*  
37: gb\_in3: \*  
38: gb\_in4: \*  
39: gb\_in5: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
C 1	280	76.5	374 14	G37241 SHGC-57295

2	45	12.3	3145	12	MM088542	U08542 Mus musculus
3	43.6	11.9	93646	7	ATF18F4	AL021637 Arabidopsis
4	41.8	11.4	119707	35	AC007180	AC007180 Drosophila
5	41.8	11.4	77685	37	AC005444	AC005444 Drosophila
6	40.4	11.0	262336	34	AC006786	AC006786 Caenorhab
7	40.4	11.0	207370	34	AC006798	AC006798 Caenorhab
8	40.4	11.0	200509	34	AC006910	AC006910 Caenorhab
9	40.4	11.0	40479	36	CEK12C11	AF043701 Caenorhab
10	39.8	10.9	104386	9	D87009	D87009 Human (Lamb
11	39.6	10.8	39565	36	CER08E3	281568 Caenorhabd1
12	38.8	10.6	145920	34	AC006710	AC006710 Caenorhab
13	38.8	10.6	299919	34	AC006712	AC006712 Caenorhab
14	38.8	10.6	110000	34	CEY11B2_0	298857 Caenorhabd
15	38.8	10.6	110000	34	CEY11B2_1	Continuation (2 of
16	38.8	10.6	110000	34	CEY11B2_2	Continuation (3 of
17	38.8	10.6	28678	36	CELT04D1	AF067617 Caenorhab
18	38.6	10.5	28992	7	SC9375	Z47071 S.cerevisia
19	38.4	10.5	177411	11	AC005696	AC005696 Homo sapi
20	38.2	10.4	97451	8	F19K23	AC000375 Sequence
21	38.2	10.4	253176	34	PFMAL13P2	AL049185 Plasmodu
22	38	10.4	13776	1	STATRSC	LI1998 Staphylococ
23	38	10.4	46445	2	AF051917	AF051917 Staphyloc
24	38	10.4	288	14	DM163A5S	Z31500 D. melanoga
25	38	10.4	34351	36	CELT14G12	U41268 Caenorhabd1
26	37.8	10.3	209365	34	AC006719	AC006719 Caenorhab
27	37.8	10.3	22846	36	CELE04A4	AF038611 Caenorhab
28	37.4	10.2	55136	9	HUMVTDHP	LI0641 Human vitam
29	37.4	10.2	9985	37	PCU43145	U43145 Plasmodu
30	37.2	10.2	37321	11	AC005498	AC005498 Homo sapi
31	37	10.1	76183	8	ATAC006570	AC006570 Arabidops
32	37	10.1	156300	11	AC005919	AC005919 Homo sapi
33	37	10.1	43410	36	CEC01G6	Z35595 Caenorhabd1
34	37	10.1	4200	37	AF145052	AF145052 Plasmodu
35	36.6	10.0	7218	5	I66494	I66494 Sequence 14
36	36.6	10.0	2157	36	TEHMIC	DI2772 Tetrahymena
37	36.6	10.0	153206	42	AC006356	AC006356 Homo sapi
38	36.4	9.9	139444	9	HS380C13	AL022161 Human DNA
39	36.4	9.9	132090	12	AC005402	AC005402 Mus muscu
40	36.4	9.9	45348	34	AC006652	AC006652 Caenorhab
41	36.4	9.9	312267	34	AC006785	AC006785 Caenorhab
42	36.4	9.9	292390	34	AC006900	AC006900 Caenorhab
43	36.4	9.9	110000	34	CEY51A2_2	Continuation (3 of
44	36.4	9.9	139259	36	CEY51A2D	AL021497 Caenorhab
45	36.4	9.9	1528	36	TETMACCAL	DI2774 Tetrahymena

## ALIGNMENTS

RESULT 1  
G37241/c  
LOCUS SHGC-57295 Human Homo sapiens STS genomic, sequence tagged site.  
DEFINITION G37241  
ACCESSION G37241  
NID 92996892  
VERSION G37241.1 GI:2996892  
KEYWORDS STS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 374)  
AUTHORS Myers, R.M.  
TITLE Human STS (1997)  
JOURNAL Unpublished (1997)  
COMMENT

Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, W-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu

Primer A: TGTGACAGATTGGACACGG  
Primer B: TGCCTTTCCCATTTGGAC  
STS size: 200  
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700

Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
Ampliflag Gold Polymerase: 0.07 units/uL  
Total Vol: 5 uL

Buffer:  
MgCl<sub>2</sub>: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

Prepared with primer pairs derived from W47380 -- Unigene.

location/Qualifiers  
1.374  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="8"  
/clone\_11b="Human"

STS  
primer\_bind 110..309  
primer\_bind 110..129  
primer\_bind complement(290..309)

BASE COUNT 131 a 59 c 74 g 110 t  
ORIGIN

Query Match Best Local Similarity 76.5%; Score 280; DB 14; Length 374;  
Matches 291; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

76 TATATTGTGTTC-AAAAAAAAAAGTCTTTGTTTAAATTACTGGTTGTGATC 134  
|||||  
374 TATATTGTGTGCAAAAAAAGTCTTTGTTTAAATTACTGGTTGTGATC 315  
|||||  
135 CATCTGCTTTTCCCATTTGGAAGTATGATTAACCATCTCGAAGTGTGAGAAAAC 194  
|||||  
314 CATCTGCTTTTCCCATTTGGAAGTATGATTAACCATCTCGAAGTGTGAGAAAAC 255  
|||||  
195 ATCTGAAGAGCTATGATCATGAGATGAGTGAATGGATGCTCAGAACATTT 254  
|||||  
254 ATCTGAAGAGCTATGATCATGAGATGAGTGAATGGATGCTCAGAACATTT 195  
|||||  
255 CACCCAGACAGCTGTTTCTATCTGTTTAAATAATTAGTTGGTCTCTACATGATA 314  
|||||  
194 CACCCAGACAGCTGTTTCTATCTGTTTAAATAATTAGTTGGTCTCTACATGATA 135  
|||||  
315 ACAAACTCTCTCAATCTGTCATATAAAGTCTGTGACTTGAAGTTAGTC 366  
|||||  
134 ACAAACTCTCTCAATCTGTCATATAAAGTCTGTGACTTGAAGTTAGTC 83

RESULT 2  
LOCUS MMU88542 3145 bp mRNA ROD 20-MAY-1997  
DEFINITION Mus musculus homeobox protein Nkx3.1 mRNA, complete cds.  
ACCESSION U88542  
NID 92105349  
VERSION U88542.1 GI:2105349  
KEYWORDS  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
MEDLINE  
97287410  
2 (bases 1 to 3145)  
Sciavolino,P.J., Abrams,E.W., Yang,L., Austenberg,L.P., Shen,M.M.  
and Abate-Shen,C.  
Tissue-specific expression of murine Nkx3.1 in the male urogenital  
system  
Dev. Dyn. 209 (1), 127-138 (1997)

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Direct Submission  
Submitted (04-FEB-1997) CABM, UMDNJ-RWJMS, 679 Hoes Lane,  
Piscataway, NJ 00854, USA  
Piscataway, NJ 00854, USA  
Location/Qualifiers  
1.3145  
/organism="Mus musculus"  
/strain="Swiss-Webster/NIH"  
/db\_xref="taxon:10090"  
14..727  
/codon\_start=1  
/product="homeobox protein Nkx3.1"  
/protein\_id="AAB58025.1"  
/db\_xref="PIR:92105350"  
/db\_xref="GI:2105350"  
/translation="MLRVAEPREPREVEAGRSPPMAAPPTGSKRLTFLIODILDRAR  
RHGSGNPDSPDRDAPEDKAGRGVAPEDEPISRPATPTPEPSDAHET  
YLDGHNPDGDAASAPDYTKOKRRAPEHSTOYIELEERSHOKYLSABRAHLAK  
NKLRTQVTKRIFQNRKTKRKQKQSEDLGVLEKNSPLSLALKDSDLSPLSYVYT  
STPYRPLTCLGSHPSFW"  
386..565  
/note="encodes the homeobox"

CDS  
misc-feature  
BASE COUNT 769 a 760 c 794 g 822 t  
ORIGIN

Query Match Best Local Similarity 12.3%; Score 45; DB 12; Length 3145;  
Matches 85; Conservative 0; Mismatches 45; Indels 3; Gaps 1;

62 ATTCTACTGAGATGATATGTTGTTCCAAAAAAGTCTTTGTTTAAATTACT 121  
|||||  
2777 ATTATAGTGTGCGATGTGTGGCAAGTCAAGAAAGGCGCTGTGTTAAATTATGT 2836  
|||||  
122 TGG---TTTGGAATCATCTCTTTTCCCATTTGGAAGTATGATTAACCATCTCT 178  
|||||  
2837 TGGGGAAGTCTGACTCAATTTTCTTTCATATCTGGAAGTATGATTAACCATCTCT 2896  
|||||  
179 GAAGTGTAGAAA 191  
|||||  
2897 GAAGTGTAGAAA 2909  
|||||

RESULT 3  
ATF18F4/c  
LOCUS ATF18F4 93646 bp DNA PLN 10-FEB-1999  
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F18F4 (ESSAII  
project).  
ACCESSION AL021637  
NID 92827644  
VERSION AL021637.1 GI:2827644  
KEYWORDS  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Capareales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS  
1 (bases 1 to 93646)  
Bevan,M., Murphy,G., Drost,L., Hall,C., Hudson,S., Ridley,P.,  
Bancroft,I., Mewes,H.W., Mayer,K. and Schueller,C.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 93646)  
EU Arabidopsis sequencing project.

TITLE  
JOURNAL

Direct Submission  
Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@dbrc.ac.uk

FEATURES  
source

```
1. .93646
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
138. .3270
/gene="F18F4.10"
complement(join(138. .251,322. .470,2694. .2855,3219. .3270))
/gene="F18F4.10"
/note="Protein sequence is in conflict with the conceptual
translation."
/codon_start=1
/product="hypothetical protein"
/protein_id="CA16599.1"
/db_xref="PID:e1248720"
/db_xref="PID:g2827645"
/db_xref="GI:2827645"
/db_xref="SPTREMBL:O49418"
/translation="MAKLCSYLEICKEVLSGFLVSSAKOLKTCSTVIKIGHPDIE
SCINCEFRVNTGPGATCRGDKYSQICTPNPVADLVKIKELNDKGLYTPISY
KLETSVRLKQEGTKEKSPSNAYADVAMILDVRLDLOEFSAGCTMLKLRACS"
complement(141. .251)
/gene="F18F4.10"
/number=1
complement(252. .321)
/gene="F18F4.10"
/number=1
complement(322. .470)
/gene="F18F4.10"
/number=2
complement(471. .2693)
/gene="F18F4.10"
/number=2
complement(2694. .2855)
/gene="F18F4.10"
/number=3
complement(2856. .3218)
/gene="F18F4.10"
/number=3
complement(3219. .3270)
/gene="F18F4.10"
/number=4
4337. .9232
/gene="F18F4.20"
complement(4337. .4424)
/gene="F18F4.20"
/number=1
complement(join(4337. .4424,4449. .4645,4739. .5174,
5945. .6138,6283. .6882,8391. .8639,8742. .9232))
/gene="F18F4.20"
/note="similarity to downy mildew resistance protein RPP5,
Arabidopsis thaliana, PATCHX:G2109275"
/codon_start=1
/product="downy mildew resistance-like protein"
/protein_id="CA16600.1"
/db_xref="PID:e1248721"
/db_xref="PID:g2827646"
/db_xref="GI:2827646"
/db_xref="SPTREMBL:O49419"
/translation="MPAFRRRPRQIYISFGNDLRKGFVSHVYKALKDARVNVFVDN
DHQMERGRPHHDDHLEVRIRHNSKALVIFESQVAESQCLNELTTHIERVAEGTL
MVIDPFRKVNIEEYNNLEGRGKCFEDMVKTOGRONQTLTHHTYGCIRSTARKEPFS
GYISNDLVLAIIIGIKKIPYISAKOTIGEESASLSAVVAALCFITPISTVP
DLGFIISPRYLAVMLLEVVVHSLRSIQORPMDPPSPPPPIPESRPRLTPPVLLT
RPRPLPYARLPPOSILPPRPQVSVSRGKELRGFVSHVYKALRINGVNFIDS
NEMKGRDLQNLFRKIEKSKALVIFSRFSESQCLNELVKIDCVKEGKLTIVPVEY
```

```
intron
/gene="F18F4.20"
/number=1
complement(449. .4645)
/gene="F18F4.20"
/number=2
complement(4646. .4738)
/gene="F18F4.20"
/number=2
complement(4739. .5174)
/gene="F18F4.20"
/number=3
complement(5175. .5944)
/gene="F18F4.20"
/number=3
complement(5945. .6138)
/gene="F18F4.20"
/number=4
complement(6139. .6282)
/gene="F18F4.20"
/number=4
complement(6283. .6882)
/gene="F18F4.20"
/number=5
complement(6883. .8390)
/gene="F18F4.20"
/number=5
complement(8391. .8639)
/gene="F18F4.20"
/number=6
complement(8640. .8741)
/gene="F18F4.20"
/number=6
complement(8742. .9232)
/gene="F18F4.20"
/number=7
12070. .13365
/gene="F18F4.30"
/number=1
12070. .13365
/gene="F18F4.30"
12070. .13365
/gene="F18F4.30"
/note="similarity to W20DMY30S"
/codon_start=1
/product="putative protein"
/protein_id="CA16601.1"
/db_xref="PID:e1248722"
/db_xref="PID:g2827647"
/db_xref="GI:2827647"
/db_xref="SPTREMBL:O49420"
/translation="MAFSFDLAVIFQERERMKRKSRYLIRRRRRDMCKSHEPMYIP
FDVIELITRLTPAKSLMRFSYKLSLSRFTFNRLRVPSFTQRLVYLTTFDNL
SLQKSKLSSSSPGSDISTSSFVDRDLTPSKMGYLSHVGLKCFKPEPVK
IYNTTROLVLPDIDIESNTIADHNNKIMYRIGHDPVGDQYKVCYIARPDDEGE
LRVYSEHWYFIIGDKSSGKRKIPCPSPHLPTTQILSLNGRMHILAWOKFDPMLVT
FDFSSSEISILQAPEDIRWFKSNPIEYGVALLNLSLRECTNLWMEDEKNNM
SEKTLVVPQMDIVKSTSLRVAGTRNNEVILVPHNRYTLTGVEVIVEPONTLLYI
FLYDLQNLARKKEIKEPYPYHTRFMDVVGGLDVENMYL"
14707. .15942
/gene="F18F4.40"
/number=1
14707. .15942
/gene="F18F4.40"
14707. .15942
/gene="F18F4.40"
/note="similarity to W15DMY30S"
```

```

/codon_start=1
/product="putative protein"
/protein_id="CA16602.1"
/db_xref="PID:e1248723"
/db_xref="PID:g2827648"
/db_xref="GI:2827648"
/db_xref="SPTREMBL:049421"
/translation="MAFSDLAIVIQERRRRRRRRDCKSRPIPEIPFIVIEIL
TRPAKSLRFRKSVKSLMSLCSRNFTNRNLKSLSPRLFMGLSSDSNHLKTVLS
LSPSPDITMSSSVTDOLTMGKGYOISHPRLMLYKSSQIYNTTRQVY
LPRIESTIAERKSKRTKTHGHPVIDQKVVCTVRASRSEVEITFLSHWVL
LEGGSRMRKISCKYPHYPLGQGLTISGRMYLAVMSDRVLVIDTSEESM
LOYGDIEMKYNGLLEGYKGLIANTYKTDIEGMELMAYVEDKMLMSKILVWPL
OLQWNSIISLGLTGRNGEVILVGPEDKTVFNILYDLQNHRIKIEIKGPDYR
LNNMEVGDMDVENIMLY"
18489..19454
/gene="F18F4.50"
/join(18489..19271,19317..19454)
/gene="F18F4.50"
/note="similarity to p48 eggshell protein gene,
Schistosoma mansoni, PATCHX:G454844"
/codon_start=1
/product="putative protein"
/protein_id="CA16603.1"
/db_xref="PID:e1248724"
/db_xref="PID:g2827649"
/db_xref="GI:2827649"
/db_xref="SPTREMBL:049422"
/translation="NDLAPBEOFLKRGILRSTSIPOYSKTFYLLITLPISE
AIIAHSLETOPILAQIDTIPADOSOLQHEWLVLFQCYILFLFAELSLAAVE
TVASLYTGKPVSSSTMSAIPVLRKLFTEFLWVSLMLAYNTAFILFVLTIVAVD
ONVLAFLVSVLFLVYVAVMLMHLASVSVLEPIYGAAKRSYELKGCYL
MCSMFEILVHCGFIAGVAVVAGGDYGFARIVAGFLSVYVYCKSPHQEI
DKSALDHLGGYGEYVPLKSNQNMENFEY"
18489..19271
/gene="F18F4.50"

Query Match      11.9% Score 43.6; DB 7; Length 93646;
Best Local Similarity 56.2%; Pred. No. 0.47;
Matches 82; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 23 AATTAAATCTTCACTGCAATTTGCAAGATTACACATTCCTGATGATATG 82
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3772 AATTAAATCTTCTAATTTATGTAAGTATATACATCATATATATGTTATG 3713

QY 83 TGTTCGAAAAAAGAGCTCTTGTAAATTAATCTTGATCCATCTGC 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3712 ATTTTTCCCATGGAAGTACATTA 168
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 TTTTCCCATGGAAGTACATTA 168
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3652 TCTTTCATATGATGATGATTA 3627

RESULT 4
AC007180/c DNA HTG 14-MAY-1999
LOCUS AC007180
DEFINITION Drosophila melanogaster chromosome 2 clone BACR01G10 (D539) RPCR-98
01.G.10 map 36D-36D strain Y2; cn bw sp, WORKING DRAFT SEQUENCE, 21
unordered pieces.
AC007180
NID 94827280
VERSION AC007180.6 GI:4827280
KEYWORDS HTG; HTGS_PHASE1;
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 119707)
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhardt, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Boyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,
Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Sult, E.,
Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Sequencing of Drosophila melanogaster
Unpublished
2 (bases 1 to 119707)
Celinker, S.E., Agbayani, A., Arcaina, T.T., Baxter, E., Blazej, R.G.,
Butenhardt, C., Champe, M., Chavez, C., Chew, M., Ciesiolka, L.,
Doyle, C.M., Farfan, D.E., Galle, R., George, R.A., Harris, N.L.,
Hoskins, R.A., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Kim, E., Lee, B., Lewis, S., Li, P., Lomtan, M.A., Mazda, P.,
Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacle, J.M., Park, S.,
Pfeiffer, B., Poon, L., Segueira, A., Sethi, H., Sult, E.,
Svitskas, R.R., Wan, K.H., Weinburg, T., Zhang, R., Zieran, L.L. and
Rubin, G.M.
Direct Submission
Submitted (29-MAR-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On May 14, 1999 this sequence version replaced gi:4764154.
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://www.fruitfly.org/sequence/) or send email
to hdg@fruitfly.berkeley.edu. All contigs in this submission meet
the following cutoffs: length >= 400 bases, phrap computed error
rate <= 1/10.
NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1043: contig of 1042 bp in length
1123: contig of 725 bp in length
1848: contig of 1275 bp in length
1928: contig of 1275 bp in length
3203: gap of unknown length
3283: gap of 676 bp in length
3959: gap of unknown length
4039: contig of 963 bp in length
5001: gap of unknown length
5081: gap of unknown length
5082: contig of 885 bp in length
5967: gap of unknown length
6046: gap of unknown length
6047: gap of 1423 bp in length
7470: gap of unknown length
7550: contig of 2057 bp in length
9607: gap of unknown length
9686: gap of 1789 bp in length
9687: contig of 1155 bp in length
11476: gap of unknown length
11556: contig of 2067 bp in length
13623: gap of unknown length
13702: gap of 3077 bp in length
16779: contig of 3077 bp in length
16859: gap of unknown length
16860: contig of 3753 bp in length
20612: gap of unknown length
20693: gap of unknown length
24724: contig of 4032 bp in length
24725: gap of unknown length
24804: gap of unknown length
29718: contig of 4914 bp in length
29798: gap of unknown length
29799: gap of 6360 bp in length
36158: contig of 6360 bp in length
36159: gap of unknown length
36238: gap of unknown length
36239: contig of 3801 bp in length
40039: gap of unknown length
40040: contig of 7225 bp in length
40120: gap of unknown length
47345: contig of 10617 bp in length
47424: gap of unknown length
47425: contig of 10617 bp in length
58241: gap of unknown length
58242: gap of unknown length
76051: contig of 17730 bp in length

```



AUTHORS	TITLE	JOURNAL	COMMENT
Waterston, R.H.	Direct Submission	Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA	On Mar 1, 1999 this sequence version replaced g1:426314.0.
			* NOTE: This is a 'working draft' sequence. It currently
			* consists of 30 contigs. The true order of the pieces
			* is not known and their order in this sequence record is
			* arbitrary. Gaps between the contigs are represented as
			* runs of N, but the exact sizes of the gaps are unknown.
			* This record will be updated with the finished sequence
			* as soon as it is available and the accession number will
			* be preserved.
	1		
	1986:	contig of 1986 bp in length	
	1987	2000: gap of unknown length	
	2001	4562: contig of 2562 bp in length	
	4563	4576: gap of unknown length	
	4577	7877: contig of 3301 bp in length	
	7878	7891: gap of unknown length	
	7892	10172: contig of 2281 bp in length	
	10173	10186: gap of unknown length	
	10187	12395: contig of 2209 bp in length	
	12396	12409: gap of unknown length	
	12410	15306: contig of 2897 bp in length	
	15307	15320: gap of unknown length	
	15321	17774: contig of 2454 bp in length	
	17775	17788: gap of unknown length	
	17789	19933: contig of 2145 bp in length	

Query Match	11.0%;	Score 40.4;	DB 34;	Length 262336;
Best Local Similarity	50.5%;	Pred. No. 2;		
Matches 98;	Conservative 0;	Mismatches 96;	Indels 0;	Gaps 0

Oy 79 ATTGCGTGCACAAAAAAGCTGCTTTGTTTAAATTACTGGTTTGTGAATCCATC 138  
|| || || || || || || || || || || || || || || || || || || || ||  
Db 80522 ATATTTGGCGGAATCAACTTAATTTTTTGAACAATTTGGCGGAATTC AAC 80581

Qy 139 TTGCTTTTCCCATGTGAACACTTAACCACTCTGAAGTGAGAAAACAATCT 198  
|| ||| | | | | | | |  
Db 80582 TTTAAATTTTGTGAAAACATTTTAGCGGAATTCAAATTTAATTTTTTTGAAAACATTT 80641

QY	199	GAAGAGCTAGTCTA	212
Db	80642	TGACGGGAATTCAA	80655

RESULT	7
AC006798/c	
LOCUS	
DEFINITION	AC006798.2_207370 bp DNA HTG 25-FEB-1999 Caenorhabditis elegans clone Y51F8, WORKING DRAFT SEQUENCE, 30 unordered pieces.
ACCSSION	AC006798
NID	94309785
VERSION	AC006798.2 GI:4309785
KEYWORDS	HTG; HTGS_PHASEI.
SOURCE	Caenorhabditis elegans . Caenorhabditis elegans . Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida; Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis 1 (bases 1 to 207370)
REFERENCE	Matereson,R.H. The sequence of Caenorhabditis elegans clone Unpublished .2 (bases 1 to 207370)
TITLE	
JOURNAL	
REFERENCE	



FEATURES	Location/Qualifiers
SOURCE	1. 207370 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /clone="Y51F8"
BASE COUNT	69298 a 36416 c 35867 g 65363 t 426 others
ORIGIN	
Query Match	11.0%; Score 40.4; DB 34; Length 207370;
Best Local Similarity	50.5%; Pred. No. 2.1;
Matches 98: Conservative	0; Mismatches 96; Indels 0; Gaps 0
Y 19 TCTGATTTAATTCCTTCAACCTTGCATTTGGCAAGATTTACACATTTCTGATGATAT 78	
DB 148972 TCAAAATTTAATTTTGTGAAAAACATTTTGGCGGGAATTCAAATTTAATTTGTTGAAA 148913	
Y 79 ATGTGTCGCAAAAAAAGAGTGTCTTGTTTAAATTAATTAAGTGTGTTGTGATCATC 138	
DB 148912 ATATTTTGGCGGGAATTCAAACTTTAATTTTGTGAAAAACATTTTGGCGGGAATTTAAAC 148853	
Y 139 TTGCTTTTCCCATTTGCACTAGTCATTACCCATCTGCACTGTAAGAAAAACATCT 198	
DB 148852 TTTAATTTTGTGAAAAACATTTAGCGGGAATTCAAATTTAATTTTGTGAAAAACATTT 148793	
Y 199 GAAGAGCTAGTCTA 212	
DB 148792 TGACGGGAATTCAA 148779	
RESULT 8	
AC006910/c	AC006910 200509 bp DNA HTG 26-FEB-1999
DEFINITION	Caenorhabditis elegans clone Y92H12, WORKING DRAFT SEQUENCE, 36
ACCESSION	AC006910
NTID	94309895
VERSION	AC006910.2 GI:4309895
KEYWORDS	HTG; HTGS_PHASE1.
SOURCE	Caenorhabditis elegans.
ORGANISM	Caenorhabditis elegans
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditidae;
	Rhabditina; Rhabditidae; Rhabditidae; Feloderinae; Caenorhabditis
	1 (bases 1 to 200509)
	Waterston, R.H.
	The sequence of Caenorhabditis elegans clone
	Unpublished
	2 (bases 1 to 200509)
	Waterston, R.H.
	Direct Submission
	Submitted (24-FEB-1999) Genome Sequencing Center, Washington
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
	MO 63108, USA
COMMENT	On Mar 1, 1999 this sequence version replaced gi:4263438.
	* NOTE: This is a 'working draft' sequence. It currently
	* consists of 36 contigs. The true order of the pieces
	* is not known and their order in this sequence record is
	* arbitrary. Gaps between the contigs are represented as
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence
	* as soon as it is available and the accession number will
	* be preserved.
	1 2013: contig of 2013 bp in length
	* 2014 2028: gap of unknown length
	* 2029 4358: contig of 2330 bp in length
	* 4359 4373: gap of unknown length
	* 4374 6822: contig of 2449 bp in length
	* 6823 6837: gap of unknown length
	* 6838 9471: contig of 2634 bp in length
	* 9472 9486: gap of unknown length
	* 9487 12251: contig of 2765 bp in length
	* 12252 12266: gap of unknown length
	* 12267 15524: contig of 3258 bp in length

Query Match	Best Local Similarity	Score	DB	Length	200509;
Matches	98; Conservative	11.0%;	50.5%;	Pred. No. 2.1;	0; Mismatches 96; Indels 0; Gaps 0;



```

gene      complement(14133..36702)
          /gene="K12C11.4"
CDS       complement(join(14133..14396,15675..16211,17274..17705,
18554..19005,19953..20502,21651..21936,22047..22244,
30994..31295,32682..33178,34086..34424,35131..35555,
36467..36702))
          /gene="K12C11.4"
          /note="contains similarity to ankyrin repeats and protein
          kinases"
          /codon_start=1
          /evidence=not_experimental
          /protein_id="AA597579.1"
          /db_xref="db:92804470"
          /db_xref="GI:2804470"
          /translation="MTSIALQQLLLPQYTLTIRSKMSKTLKNGSPFVYFSAC
SPSKSCISGNTTFRFLYNALPLFIARORRIGOPAVRRVRKRGKAAKFI
KRRVATSRGRTRONIERVRYLQKIRGNSVNEVLAHYETASDIYDELVSSEL
FDHVCAKECLDEVEAAAFIKQILLAVRHLSLHVLDIKPENVMKQKQDSQIKIID
FGLSREIEGAVVAGTPEEVAPEVVAEYALSPATDMVAENVVYIILSGSPFLG
DNDETFSTIRVRYHFSDRYEKTSKHADFIYRLFVDDVRATEVEELQHPWIRG
PEGNAIDIRKASCITTSIOSFKTRORMKRCVELVAVLKASSSRIRIDGRDEEDM
VASCTLCAEGCNLRALKLSALHKLIPNATRSKLSSESEPGATAMCAATYGAIE
VRYTFMKGNCITARDNGDTPLVACRFQHTVAGIVANERKIDVDSINKTGETALHC
AVESADTRVRLQLRLPLDPLNAREEPLHVAAGVHAGVQALLDLNAPIDAVEQ
DKETALILENGNVDAISILITNGCDINADHGGDTALHASKGLDAVOTLCHCA
VYDVSANANKETALHLAAGHYDIIKVLRAADVTLRDDGLTAELVAVERLEA
HSLIKVKSQEIREEYISOLYPLDTSIRIKKLIGHOSGKRLVOTLHSSGTSIF
LESVTRPIDHYSPSSSMKDDQIHSTNGSVSSNNSSPDLLAAAGSKRAPHSTYT
RGIDVTVINNGGERSVMEFGIEPMHTCTDFVGNADCTILILRTSDPEVQKQ
KLTVMFLGRVTPPEPIGHGCFSSSRSKVIYGTHTASLSLPOMNQBEGYSSDIEA
MLNTVLRREFTHMDHRLILDATNPSCGMKTLKEMELAKRTNILLAKTLALID
TYVNHNLVKKHANEPTVTPPEIOLVREINELISGADAROIYOOLDIGLVYLR
NDICDADYVIVANEMEGTHTIGQLLSAEPLSKSPNGSVTSSALTFPIEPQSLM
TIEVYQLCAPDARTCAHEPPVITOTEPADSTIRPYSLEKEKDYTGVRRIIPMGIM
ERSLHSTFPIQVALRSINDVQPAKDTQHLQWSESKVSDREAVIRKVGAVAIR
ARGPESATSMFYFEMDLINLYEHAAEVGPGLSRHFISSPKHLEKREHHPALFPEE
SMEMOORELSYKGTODEBELFTDYVCGESRVSARHLTIGIDVADLOMAERCELA
CLIDPHAMGRDMSILAVKQLTDOVDVDSGTOSRTDQLNEMAHHPEDGASVGN
LCRIIVEIGRCRDALRYRPLYFAPLEDQFLFTNDGVSCHSSSEHNPINI"
          repeat_region
          /note="C. elegans telomere-like hexamer repeat"
BASE COUNT 13418 a 7528 c 7289 g 12244 t
ORIGIN
Query Match      11.08; Score 40.4; DB 36; Length 40479;
Best Local Similarity 50.58; Pred. No. 2.8;
Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;
OY 19 TCTGAATTTAATCTTCACTTGCAGATTCACATTCACATTCAGTGTAT 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24634 TCAATTATTATTTTGGAAACATTTGGCGGAATTCAAATTTAATGTTTGA 24575
OY 79 ATTGCTTGCAMAAAAAAGTCTCTGTTTAAATTAATCTGTTTGAATCCATC 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24574 ATATTGGCGGAATTCAAACTTAATTTTGGAAACATTTGGCGGAATTC 24515
OY 139 TTGCTTTTCCCACTTGGAACTAGTCAATTAACCATCTCTGAACTGGTAGAAA 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24514 TTTAATTTTGGAAACATTTTAGCGGAATTCAAATTTAATTTTGGAAACAT 24455
OY 199 GAAGAGCTAGCTA 212
DB 24454 TGACGGGAATTCAA 24441

```

```

KEYWORDS  3'Oxyl.1; 5'Oxyl.1; V3-1P; V1-11; V4-1; V1-12P; immunoglobulin
          light chain.
SOURCE    Homo sapiens human pre-pro B cell line cell line:FLB14-14 DNA,
          clone_1ib:total human BAC library; Kelo BAC library clone:288A10.
ORGANISM  Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 104386)
AUTHORS   Shimizu,N.
TITLE     Direct Submission
JOURNAL   Submitted (08-AUG-1996) to the DDBJ/EMBL/GenBank databases.
          Nobuyoshi Shimizu, Kelo University School of Medicine, Depart-
          ment of Molecular Biology, 35 Shinanomachi, Shinjuku-ku, Tokyo 160,
          Japan (E-mail:shimizu@db.med.kelo.ac.jp, Tel:03-3351-2370,
          Fax:03-3351-2370)
REFERENCE 2 (sites)
AUTHORS   Kawasaki,K., Minoshima,S., Nakato,E., Shibuya,K., Shintani,A.,
          Schmeits,J.L., Wang,J. and Shimizu,N.
TITLE     One-megabase sequence analysis of the human immunoglobulin lambda
          gene locus
JOURNAL   Genome Res. 7 (3), 250-261 (1997)
FEATURES  Location/Qualifiers
           1..104386
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /cell_line="FLB14-14"
             /cell_type="human pre-pro B cell line"
             /chromosome="22"
             /clone="288A10"
             /clone_1ib="total human BAC library; Kelo BAC library"
             /map="22x11.2"
           1..104386
             /note="region III"
             /gene="V1-12P"
             /gene="V1-12P"
             /gene="V1-12P"
             /gene="V1-12P"
             /pseudo
             /codon_start=1
             /codon="8827..9138"
             /gene="V4-1"
             /gene="V4-1"
             /gene="V4-1"
             /codon_start=1
             /protein_id="BAA20003.1"
             /db_xref="PID:d1020804"
             /db_xref="PID:g2114253"
             /translation="QPVLTQPPSSASPGESARLCTLPSPDINSGSYNYWQKPGS
PRRLYLYSDSDKGGSGVPSRFSGSKDASANTGILLISGLQSEDEADYICAMDPNSN
AS"
             13275..13566
               /gene="V1-11"
               <13275..>13566
               /gene="V1-11"
             /codon_start=1
             /protein_id="BAA20004.1"
             /db_xref="PID:d1020805"
             /db_xref="PID:g2114254"
             /db_xref="GI:2114254"
             /translation="QGVLTQPPSVSEAPRQRTYISGSGSSNIGNNAVWVYQOLPEKA
FKLLIYDLPSPGVSDFSGSKSGTSASLSIGLQSEDEADYICAMDPNSLN"
             31459..31753
               /gene="V3-1P"
               /gene="V3-1P"
               /gene="V3-1P"
               /pseudo
             /codon_start=1
             /complement(68859..70486)
             /gene="5'Oxyl.1"
             /complement(68859..>70486)
             /gene="5'Oxyl.1"

```

```

RESULT 10
LOCUS    DB7009      104386 bp      DNA      PRI      07-FEB-1999
DEFINITION DB7009 Human (lambda) DNA for immunogloblin light chain.
ACCESSION DB7009
NID      g2114252
VERSION  DB7009.1 GI:2114252

```

```

/note="putative"
/codon_start=3
/protein_id="BAA20005.1"
/db_xref="PID:d1020806"
/db_xref="PID:g2114255"
/db_xref="GI:2114255"
/translation="QSCCEKEPEPQKNIQETKOYDEDAELLIFVGEVHNEDALLIF
VQVSNKSPVSNILNRLTSGMSRRKRYDLRKDTARKLOPKSHETVTSAAVVLPA
SLESRTSDSPDILEPLSKPDYRNSPQVPPNNSSELPSTFETDLSHHPSTALYS
GGINSPRVSQKSTEEVNSINPKRAKLRDGIENSSASPSDEPHMTMOQSTPSN
NHTSLSHVONGAPPPAPFPRKNDIHKRINTNLDRENELATKTDLSLSQKRTDPK
ENPIVLSDFYGGHKGSGQPEOKHTTFKCLSCVKYLNKRNKRNHVAHLEEKORR
DSWENHTTCQCHROFPFPDQCHIEVHTAQPSTVCKICELSEFETDVLQHMKD
HHKPEMPYVOYCHYRNSVADQVETHTRCHENTKLLCEFCIKLFEKTAAPYCHYR
GHMKSAHOCSCKRLQFLFKEMKHTKQCHOMEKKPKQLGELPETKVTIQTSLLEPL
QPGSYDVASITVTSDESPSLPRSKSKSKSH"
complement(95093..96717)
/gene="3'0Y11.1".>96717)
/complement(95093..>96717)
/gene="3'0Y11.1"
/note="putative"
/codon_start=3
/protein_id="BAA20006.1"
/db_xref="PID:d1020807"
/db_xref="PID:g2114256"
/db_xref="GI:2114256"
/translation="DIFLCRRVSPKRLRESKQREDEDEPDILYGVGEVHNDAEY
LEVGMISNRKPVSNILNRLTSGMSRRKRGHFRQYPAHVSOPLANHTSMKAIAPY
LSEGRSDSPYTKMSSSEPKMSPOVSPNYSLSLPRGQCLVGAIVSGGRRNS
PDKRLSTSDINSRDSKRVKLDGIPVPSLAVPSDMSSTITSTVPSOGICNSNH
ONGVTFPWPDANGAHFNLTPERANESGLAMTDISSLASQKTFDPKKNPVIYLSN
FYGOHKGDGQPEOKHTTFKCLSCVKYLNKRNKRNHLEFEKORNDSEMTTC
OHCHROFPFPDQCHIEVHTAQPSTVCKICELSEFETDVLQHMKDHHKPEMPY
VCQCHYRNSVADQVETHTRCHENTKLLCEFCIKLFEKTAAPYCHYRNSVADQ
CSKRLQFLFKEMKHTKQCHOMEKKPKQLGELPETKVTIQTSLLEPLQPGSYDV
SNTPDQSPVTKKTAANTKIDSRLEPKSDSS"

```

```

BASE COUNT      31195 a 22463 c 22062 g 28666 t
ORIGIN
Query Match      10 9%: Score 39.8; DB 9; Length 104386;
Best Local Similarity 57.7%: Pred. No. 3.2;
Matches 71: Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```

```

OY 5 TTTCCTCCCCCTCTGTAATTTTCACTGCAATTCGAAGATTACACATT 64
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 51580 TTCTCTCTCTCTGCTGAAATCTCTTTATCTGACAAATTATACCCCTGTAT 51521
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 65 TCACGTGTATATATGTTGTCAAAAAAGTGTCTTTGTTTAAATTAATTG 124
| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 51520 TTCTCTGACGATAGAGATTGAGAAATATATTAATTCTTATTTAGTTT 51461
||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 125 TTT 127
| |
DB 51460 ATT 51458

```

```

RESULT 11
CEK08E3      39565 bp      DNA      INV      23-NOV-1998
DEFINITION  Caenorhabditis elegans cosmid K08E3, complete sequence.
ACCESSION   Z81568
NID         93643297
VERSION     Z81568.1 GI:3643297
KEYWORDS    HTG.
SOURCE      Caenorhabditis elegans.
ORGANISM    Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilitida;
Rhabdilitina; Rhabdilitoidea; Rhabdilitidae; Peloderinae; Caenorhabditis.
REFERENCE   1 (bases 1 to 39565)
AUTHORS    McMurtry, A.
TITLE      Direct Submission
JOURNAL    Submitted (06-NOV-1996) Louis, MO 63110, USA. E-mail:

```

```

REFERENCE
AUTHORS      Jee@sanger.ac.uk or iw@nematode.wustl.edu
2 (bases 1 to 39565)
WILSON, R., Almsough, R., Anderson, K., Baynes, C., Berks, M.,
Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J.,
Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello, A.,
Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M.,
Johnston, L., Jones, M., Kersey, J., Kirsten, J., Laiter, N.,
Latreille, P., Lighthouse, J., Lloyd, C., McMurtry, A., Mortimore, B.,
O'Callaghan, M., Parsons, J., Percy, C., Riffen, L., Roopra, A.,
Saunders, D., Showkeen, R., Smalton, N., Smith, A., Sonhammer, E.,
Studen, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,
Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,
Wilkinson-Spoat, J. and Mohlman, P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
MEDLINE
COMMENT
TITLE
FEATURES
source
location/Qualifiers
1..39565
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="K08E3"
3510..6358
/gene="K08E3.1"
join(3510..3749,3844..4023,4219..4319,4374..4816,
4998..5101,5288..5683,5730..6167,6215..6358)
/gene="K08E3.1"
/note="similar to common central domain of tyrosinase;
cDNA EST EMBL:Z14371 comes from this gene; cDNA EST
CEMSG81F comes from this gene; cDNA EST EMBL:M89185 comes
from this gene; cDNA EST EMBL:C13572 comes from this gene;
cDNA EST yk240a11.3 comes from this gene; cDNA EST
yk240a11.5 comes from this gene"
/codon_start=1
/protein_id="CA04594.1"
/db_xref="PID:e1348001"
/db_xref="PID:g3878440"
/db_xref="GI:3878440"
/translation="MRWPTLITLTLCLLVLLISVAGVSTTKGAVGATVARKTVK
VAVKNTIORMNOKRMRPEPRDCSDAPKHKIKOTCIIMRMDKATRRRLRAGVLRNP
NOVDNNTQPIPVANRAGOAHYPRDCMTLCLCFPFNGNNGVQCCLSVGLYNLMSW
RKETRMKTEERRRMHNLTLKRNGETIDLSQSHLEGVGSGHSGPGLLPVHRETL
KRVEIALRWVDVTFIPYWDVMSDYLDPDRDSIMPSDLEVGDTGYGVNVTGPFAVY
RTTEGRSTILRNLAGEQLEFNEQVNTIVAFQNTIENTLATYAPQPCPYNNNGALEY
TSHNIIHLMIGDKDQPESTANEPFEHMHSEFVLYELMQLQOPMLREOASADHP
TCANWHEFSYAPRPPEYLVNRDGLSYSDQWYRVARPATGSHORNGSPYLPDDT
RGYPHCYSKRLNCGSGFQNDACVASRCRMRCVNAFNAKMKARTANSITLIDL
ELAKESVNGTKEVRYPLVAKISIKMTAPLFVDCYNMRPCDSWANNGCQREPEHS
OYCAQSCNTQPSYNTDCAKMDRHNVCVAKAKDKMSGKSADPMSCRSKSCSIT
RESOCFNNKRIIVPSONRESISDSLMSEVTNRIRIKYKVAKEEVEESTKKVRHN"

```

gene  
CDS  
gene  
gene  
CDS  
CDS  
gene  
gene

```
complement(6755..7969)
/gene="K08E3.2"
complement(join(6755..6953,7015..7141,7190..7388,
7618..7680,7727..7969))
/gene="K08E3.2"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB04590.1"
/db_xref="PID:e1347997"
/db_xref="PID:g3878436"
/db_xref="GI:3878436"
/translation="MEYEVLYALISLCSLIIGCOOTSKRAQSVAPIMKEHAESITR
TPVKSAPAVSLSLALEDDEYDEKKREGESEEESEKREDEEKEDKSDSEKDD
EDHEKKRTAEKENNEKKDENKNKKDKDDEKDDKODKODEDENNDKKEKKEDOK
DDRKAASMSVDAPELIFERTITSQRKLKLKLTKLRFAKFKICSSVWYIINPFGK
IDPMSTAEVTITHRSEKODEKLVIVPAKMKEKEIEMAAQAFQIKKTGSALTVLVSS
"
14167..18631
/gene="K08E3.3a"
14167..18631
/gene="K08E3.3b"
join(14167..14295,14596..14763,14810..14911,15156..15314,
15357..15472,15522..15645,15694..15854,15902..16007,
16055..16259,16305..16386,16433..16509,16656..16905,
17201..17352,17487..17593,17642..17692,18137..18266,
18350..18534,18568..18631)
/gene="K08E3.3b"
/note="similar to EF hand, Src homology domain 3; cDNA EST
yk273e3.3 comes from this gene"
/codon_start=1
/protein_id="CAB04595.1"
/db_xref="PID:e1348002"
/db_xref="PID:g3878441"
/db_xref="GI:3878441"
/translation="MPEKAVSTCFHVTLTKDYIVFYVSNLNLPFASRMPRRRGRLT
PALRAMNLNPEALLAEYSKARTSGARTVEEPIPKKYEAAGCRSTTLKEPTLL
SHFDREGNQIHLMLKSLMQNGEKMTQENRALKPATGDTITRNLPNKHFTLDH
VAIGLEIRTVQVOQNIIPVSREFIVQDPNSALLSTOGIDFEARLOFSKAKIAIE
ESTKRSLAKRYARKSEEDILKSVSTSFNSFLQOOLDIARHOTSAKHIGGV
VYVASVKTCOMSSSRNAINDKTTINDKLEDOINEMCKSGKCYLFSPDAEYSYKFY
KADNMETIRLEEKARALANARNKEACELADQYSAVLRKTAPOKRVHLLPVIFA
RKANDKEIDAMROYLOKTYFSDSLADSTECKINMOREKVAKIDAEGDAQLYKSV
EATIEDPAPELTEDGDPRNCDSRTNDSADSGGLVLSLSSPSKNNIINFGLIKE
ADEKPEASNNOIDMDKSCPAHVRSLCLRSIRIMEKOLOAOIGREGIRILOQAY
TNPOHQNPACTEPLISYAKRIEKLMDIHNLKEFYAMLEMSVEGORSRGGRPD
TRMSGNSTNONSSKTIEDVSGEAGNSSADSSKNILROLFTTPTRLISSPTSK
SSTPPLARRAIESPKTIRSFSGAIRCSISTDPSVYETRAYVTALFEKASAEI
MSIDEGETILYLIEHDHGDCWTTRKCKRKNESGCVPPSYIDFPQ"
join(14167..14295,14596..14763,14810..14911,15156..15314,
15357..15472,15522..15645,15694..15854,15902..16007,
16055..16259,16305..16386,16433..16509,16656..16905,
17201..17352,17487..17593,17642..17692,18137..18266,
18350..18534,18574..18631)
/gene="K08E3.3b"
/note="predicted using GeneFinder; similar to EF hand, Src
homology domain 3; cDNA EST EMBL:D64948 comes from this
gene; cDNA EST EMBL:D66852 comes from this gene; cDNA EST
EMBL:D76408 comes from this gene; cDNA EST EMBL:D70789
comes from this gene; cDNA EST EMBL:D68217 comes from this
gene; cDNA EST EMBL:C13458 comes from this gene; cDNA EST
yk273e3.3 comes from this gene"
/codon_start=1
/protein_id="CAB04591.1"
/db_xref="PID:g347998"
/db_xref="PID:e1347998"
/db_xref="PID:g3878437"
/db_xref="GI:3878437"
/translation="MFEKAVSTCFHVTLTKDYIVFYVSNLNLPFASRMPRRRGRLT
PALRAMNLNPEALLAEYSKARTSGARTVEEPIPKKYEAAGCRSTTLKEPTLL
SHFDREGNQIHLMLKSLMQNGEKMTQENRALKPATGDTITRNLPNKHFTLDH
VAIGLEIRTVQVOQNIIPVSREFIVQDPNSALLSTOGIDFEARLOFSKAKIAIE
ESTKRSLAKRYARKSEEDILKSVSTSFNSFLQOOLDIARHOTSAKHIGGV
VYVASVKTCOMSSSRNAINDKTTINDKLEDOINEMCKSGKCYLFSPDAEYSYKFY
KADNMETIRLEEKARALANARNKEACELADQYSAVLRKTAPOKRVHLLPVIFA
RKANDKEIDAMROYLOKTYFSDSLADSTECKINMOREKVAKIDAEGDAQLYKSV
EATIEDPAPELTEDGDPRNCDSRTNDSADSGGLVLSLSSPSKNNIINFGLIKE
```

```

gene
  complement(join(18904..19472,19520..20118,20164..20373,
    20433..20734,21052..21164,21214..21352))
  cds
    /note="predicted using GeneFinder: similar to
    Col11n/tropomyosin-type actin-binding proteins, Src
    homology domain 3; cDNA EST yk442b7.5 comes from this
    gene; cDNA EST EMBL:Z14844 comes from this gene; cDNA EST
    EMBL:W88991 comes from this gene; cDNA EST EMBL:Z14502
    comes from this gene; cDNA EST EMBL:D71424 comes from this
    gene; cDNA EST EMBL:D74120 comes from this gene; cDNA EST
    EMBL:D71594 comes from this gene; cDNA EST EMBL:D74252
    comes from this gene; cDNA EST EMBL:J72634 comes from this
    gene; cDNA EST EMBL:D68488 comes from this gene; cDNA EST
    yk207d8.3 comes from this gene; cDNA EST yk207d8.5 comes
    from this gene; cDNA EST yk274h9.3 comes from this gene;
    cDNA EST yk274h9.5 comes from this gene; cDNA EST
    yk338d11.3 comes from this gene; cDNA EST yk338d11.5 comes
    from this gene; cDNA EST yk335b5.3 comes from this gene;
    cDNA EST yk335b5.5 comes from this gene; cDNA EST
    yk339a4.3 comes from this gene; cDNA EST
    yk339a4.5 comes from this gene; cDNA EST yk376a8.3 comes
    from this gene; cDNA EST yk376a8.5 comes from this gene;
    cDNA EST yk376a8.5 comes from this gene; cDNA EST
    yk339a9.3 comes from this gene; cDNA EST yk339a9.5 comes
    from this gene; cDNA EST yk419c7.3 comes from this gene;
    cDNA EST yk442b7.3 comes from this gene; cDNA EST
    yk462c7.3 comes from this gene; cDNA EST yk462c7.5 comes
    from this gene; cDNA EST yk465d12.3 comes from this gene;
    cDNA EST yk465d12.5 comes from this gene"
    /codon_start=1
    /protein_id="CAB04592.1"
    /db_xref="PID:el347999"

Query Match      10.8% Score 39.6 DB 36 Length 39565:
Best Local Similarity 51.1% Pred. No. 4.2:
Matches 93: Conservative 0: Mismatches 89: Indels 0: Gaps 0:

Cy 23 AATTATTAATCTTCACTGCAATTTGCAAGATACACATTTCACTGATGATATTC 82
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8505 AATTCAAATTTCTGGAATCATTTTGGCGGGAATTCAAATTTAATTTTAAAAAAT 8564
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Cy 83 TGTTCGCAAAAAAAGAGTGTCTTTGTTAAATTTACTTGTTGTAATCCATCTTGC 142
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8565 TTTGGCGGGAATTAATAATTTTAAATTTTGAATAATTTTGGCGGGAATTTAAATTTT 8624
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Cy 143 TTTTCCCATTTGGAAGTACTGATTAACCACTCTCGACTGTGTAGAAAAACATCTGAAG 202
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 8625 AATTTCTGAAAAATTTTGGCGGAATTCAAATTTTAATTTTGAATAATTTTGGCG 8684
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Cy 203 AG 204
Db 8685 GC 8686

RESULT 12
AC006710 AC006710 145920 bp DNA HTG 23-FEB-1999
LOCUS Caenorhabditis elegans clone Y119C1AA, WORKING DRAFT SEQUENCE, 1
DEFINITION
unordered pieces.
ACCESSION AC006710
NID 94263228
VERSION AC006710.1 GI:4263228
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

```





---



1	366	100.0	366	1	V58588	Prostate tumour sp
2	366	100.0	366	1	V61203	CDNA sequence of p
3	265	72.4	335	1	V58614	Prostate tumour sp
4	265	72.4	335	1	V61229	CDNA sequence of p
5	38	10.4	10322	1	V74409	Staphylococcus au
6	34	9.3	3012	1	Q21188	Sequence encoding
7	33.6	9.2	4483	1	O03253	Sequence of clone
8	33.6	9.2	4483	1	N00843	Tobacco basic beta
9	33.6	9.2	4483	1	V72993	Tobacco basic beta
10	33.6	9.2	4483	1	V62789	Tobacco beta-1.3-
11	33.4	9.1	3601	1	V74317	Staphylococcus au
12	32.2	8.8	110000	1	X20248_08	Continuation (9 o
13	32.2	8.8	10715	1	X20248_09	Continuation (10 c
14	32	8.7	40352	1	V02032	MAGE-B cluster DN
15	32	8.7	1407	1	X16568	Protein phosphatase
16	32	8.7	110000	1	X20248_04	Continuation (5 o
17	32	8.7	116277	1	X20249	Borrelia burgdorferi
18	31.6	8.6	3504	1	O42565	RTS expression cas
19	31.6	8.6	3504	1	O42567	RTS* expression ca
20	31.6	8.6	20072	1	X13026	Enterococcus faec
21	31.4	8.6	10502	1	X20258	Borrelia burgdorferi
22	31	8.5	1593	1	O66903	P. falciparum prot
23	31	8.5	445	1	V86233	EST clone S169. Ne
24	31	8.5	110000	1	X20248_06	Continuation (7 o
25	31	8.5	111309	1	X20250	Borrelia burgdorferi
26	30.8	8.4	14180	1	V05159	Mouse butyrophillin
27	30.8	8.4	58407	1	V21210	Methanococcus jann
28	30.8	8.4	1322	1	X13419	Enterococcus faec
29	30.8	8.4	211	1	X13816	Enterococcus faec
30	30.6	8.4	748	1	V74609	Staphylococcus au
31	30.6	8.4	110000	1	X20248_07	Continuation (8 o
32	30.6	8.4	5002	1	X25011	Antibiotic-specific
33	30.4	8.3	3755	1	O31890	EMR-1 gene for det
34	30.4	8.3	110000	1	T58840_0	Mycoplasma genital
35	30.4	8.3	1875	1	V14014	Xenopus "Irazzled"
36	30.4	8.3	12445	1	X13119	Enterococcus faec
37	30.2	8.3	5336	1	T90532	Arabidopsis male f
38	30.2	8.3	2127	1	T96839	Intron 9 of human
39	30.2	8.3	9399	1	X20261	Borrelia burgdorferi
40	30	8.2	2960	1	T41616	Adzuki bean endo-
41	30	8.2	11465	1	V74755	Staphylococcus au
42	29.8	8.1	3083	1	O11851	Glutamate receptor
43	29.8	8.1	110000	1	X20248_02	Continuation (3 o

DD 301 11AG1C 300

```
RESULT 2
V61203
ID V61203 standard; cDNA; 366 BP.
AC V61203;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.

PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PR used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 85; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 366 BP; 102 A; 78 C; 58 G; 128 T;

Query Match 100.0%; Score 366; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 9.2e-90;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCTCTTCTCCCTCCCTCCGTAATTATCTTCAATTCATTGCAAGATTACA 60
DB 1 GCTCTTCTCTCCCTCCCTCGTGAATTTATCTTCAATTCGCAAGATTACA 60
OY 61 CATTTCACGTGATGATATGTTGTGGCAAAAAAAGTGTCTTTGTTAAATTAC 120
DB 61 CATTTCACGTGATGATATGTTGTGGCAAAAAAAGTGTCTTTGTTAAATTAC 120
OY 121 TTGTTTGTGAATTCATCTTGTCTTTTCCCATTTGGAATGATCTGTA 180
DB 121 TTGTTTGTGAATTCATCTTGTCTTTTCCCATTTGGAATGATCTGTA 180
OY 181 ACTGGTAGAAAAACATCTGAAGAGCTAGTCTATCAGCATCTACAGGTGAATTGGATGTT 240
DB 181 ACTGGTAGAAAAACATCTGAAGAGCTAGTCTATCAGCATCTACAGGTGAATTGGATGTT 240
OY 241 TCTCAGAACCATTTACCCAGACAGCTGTTCTATCTGTTTAATTAATTAGTTGGT 300
DB 241 TCTCAGAACCATTTACCCAGACAGCTGTTCTATCTGTTTAATTAATTAGTTGGT 300
OY 301 TCTCTACATGATGATTAACAACCTGTCTCCATGTCATCAATAAAGTCTGTGACTGAACT 360
DB 301 TCTCTACATGATGATTAACAACCTGTCTCCATGTCATCAATAAAGTCTGTGACTGAACT 360
OY 361 TTAGTC 366
DB 361 TTAGTC 366

RESULT 3
V58614/c
ID V58614 standard; cDNA; 335 BP.
AC V58614;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
PN M09837418-A2.
PD 27-AUG-1998.
```

```
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PR for detecting and treating prostate cancers
PS Claim 1; Page 100; 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 335 BP; 117 A; 55 C; 66 G; 94 T;

Query Match 72.4%; Score 265; DB 1; Length 335;
Best Local Similarity 98.9%; Pred. No. 1.2e-62;
Matches 265; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 99 AAGTGTCTTGTAAATTAATCTGTTGGTGGATGCATCTGCTTTTCCCATGGAA 158
DB 335 AAGTGTCTTGTAAATTAATCTGTTGGTGGATGCATCTGCTTTTCCCATGGAA 158
OY 159 CTAGCATTAACCCATCTGTAAGTGTAGAAAAACATCTGAAGAGTATGATCAGCA 218
DB 275 NTAGCATTAACCAATCTGTAAGTGTAGAAAAACATCTGAAGAGTATGATCAGCA 216
OY 219 TCTGACAGGTGAATGATGTTCTCAGAACCATTTACCCAGACGCTGTTCTATCC 278
DB 215 TCTGACAGGTGAATGATGTTCTCAGAACCATTTACCCAGACGCTGTTCTATCC 156
OY 279 TGTTAATAAATTTGTTGGTCTCTACATGATCAACAAACCTGCTCAATGTGCAC 338
DB 155 TGTTAATAAATTTGTTGGTCTCTACATGATCAACAAACCTGCTCAATGTGCAC 96
OY 339 ATAAAGCTGTGACTTGAAGTTAGTC 366
DB 95 ATAAAGCTGTGACTTGAAGTTAGTC 68

RESULT 4
V61229/c
ID V61229 standard; cDNA; 335 BP.
AC V61229;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.

PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PR used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 94; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
```



```

FT      /*tag= e
PN      WO9201792-A.
PD      06-FEB-1992.
PE      24-JUL-1991; F00607.
PR      24-JUL-1990; FR-009460.
PA      (SNEI) SANOFI SA.
PA      (ERAP) SOC NAT ELF AOUTTAINE.
PI      Dubois M, Grlson R, Leguay JJ, Pignard A, Toppan A;
DR      WPI: 92-064950/08.
DR      P-PEDB: R20820.
PT      New recombinant protein with endo chitinase activity - also DNA
PT      encoding it and plants which express it, resistant to fungi,
PT      insects, bacteria and nematodes.
PS      Example: Fig 2: 82pp; French.
CC      The inventors claim a new recombinant gene which codes for a protein
CC      with endochitinase activity, or its precursor (see Q21007). The
CC      coding part of the gene contains at least the 5' part of DNA for
CC      tomato endochitinase and at least the 3' part of DNA for the tobacco
CC      enzyme (DNA can be genomic or cDNA, but at least one intron is pref.
CC      present). The recombinant gene includes the 35S promoter of
CC      cauliflower mosaic virus plus the terminator from the nopaline
CC      synthase gene of Agrobacterium tumefaciens.
SQ      Sequence 3012 BP; 965 A; 634 C; 436 G; 977 T;

Query Match
Best Local Similarity 48.9%; Score 34; DB 1; Length 3012;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY      10 CTCCTCTCTGATTTATTTCTTCAACTTGCAATTTGCAAGATTACACATTTCAC 69
DB      1 ||||| ||| ||| ||||| ||| ||| ||||| |||||
DB      405 CCCCCCCCCCCCCCAAAAATATTTAAATTTGTTTAAAAATTTTCAATTTCAAA 464
QY      70 GGTATGTATGTGTGTCGCAAAAAAAGGCTTGTGTTAAATACTGTTTCT 129
DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      465 ATTTTATTTCTTACTTAAATAAAGATATATTCACAAAATTTTACTTACTT 524
QY      130 GAATCCATCTGCTTTTCCCATTTGGAAGTACTTAACCATCTCGAATGCTAGA 189
DB      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB      525 CAAAACCAACACACTTTTCCGAAAAAATTTCTATTCCACCAACCAATATGAGAAAT 584
QY      190 AAAACA 195
DB      |||||
DB      585 AATCA 590

RESULT 7
Q03253/c
AC      003253 standard; DNA; 4483 BP.
ID      003253;
DE      25-JUL-1990 (first entry)
DE      Sequence of clone pBSgluc39.1 encoding beta-1,3-glucanase from genomic
DE      sequence.
KW      Beta-1,3-glucanase; papermaking; polysaccharide hydrolase;
KW      ds.
OS      Nicotiana sp.
PN      EP-353191-A.
PD      31-JAN-1990.
PE      20-JUL-1988; 810555.
PR      29-JUL-1988; US-353312.
PA      (CIBA) Ciba Geigy AG.
PI      Shlunshi H, Wenzler H, Hofsteenge J, Ryals J, Sperisen C;
DR      WPI: 90-031717/05.
PT      Recombinant DNA encoding beta-1,3-glucanase polypeptide(s) -
PT      useful in plant protection and biomass conversion pref. isolated
PT      from tobacco.
PS      Disclosure; Fig 8: 41pp; English.
CC      Beta-1,3-glucanase is involved in plant defence against pathogens and as
CC      it hydrolyses polysaccharides, useful in biomass conversion or paper
CC      making. The clones allow for transfer of the gene to other species and
CC      control of gene expression.
SQ      Sequence 4483 BP; 1418 A; 838 C; 744 G; 1483 T;

```

```

Query Match
Best Local Similarity 52.1%; Score 33.6; DB 1; Length 4483;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY      1 GCTCTTCTCTCCCTCCCTCGAATTTAATTTCTTCACTTGCAATTTGCAAGATTACA 60
DB      4481 GCTCTTCTCTCTCTCTCGAATAGAGTCTTCTTCTCTGCTTGGAGTGAATGAGA 4422
QY      61 CATTTCACTGATGATATATGTGTGTCGCAAAAAAAGTGTCTTGTGTTAAATTAC 120
DB      4421 CATTTCAACGAGAGCTAGTTGTGGTAAGATCAAAAATGATCTTGTAGCTGGAAGTTAT 4362
QY      121 TTGTTTGTGAATCCATCTGCTT 144
DB      4361 CCGGTTTGTGTCATGCTTCTT 4338

RESULT 8
N90843/c
ID      N90843 standard; cDNA; 4483 BP.
AC      N90843;
DE      24-JAN-1990 (first entry)
DE      Tobacco basic beta-1,3 glucanase gene.
DE      Tobacco basic beta-1,3 glucanase; pBSgluc39.3; herbicide resistance.
OS      Nicotiana tabacum cv. Xanthi.
PN      EP-332104-A.
PD      13-SEP-1989.
PE      06-MAR-1989; 103888.
PR      08-MAR-1988; US-165667.
PA      (CIBA) Ciba Geigy AG.
PI      Ryals J, Montoya A, Harms C, Duesing J, Sperisen C, Melns F,
PI      Payne G;
DR      WPI: 88-265342/37.
PT      Chemical induction of cloned genes in plants - DNA sequence and vectors
PT      and processes for inducible expression leading to herbicide and insect
PT      resistance.
PS      Claim 79: page 72-74; 117pp; English.
CC      Sequence is a chemically inducible gene used to regulate
CC      transcription of an associated DNA sequence in plant
CC      tissue.
CC      See also N90368, N90841-N90846, and P91384-P91386.
SQ      Sequence 4483 BP; 1418 A; 837 C; 745 G; 1483 T;

Query Match
Best Local Similarity 52.1%; Score 33.6; DB 1; Length 4483;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY      1 GCTCTTCTCTCCCTCCCTCGAATTTAATTTCTTCACTTGCAATTTGCAAGATTACA 60
DB      4481 GCTCTTCTCTCTCTCTCGAATAGAGTCTTCTTCTCTGCTTGGAGTGAATGAGA 4422
QY      61 CATTTCACTGATGATATATGTGTGTCGCAAAAAAAGTGTCTTGTGTTAAATTAC 120
DB      4421 CATTTCAACGAGAGCTAGTTGTGGTAAGATCAAAAATGATCTTGTAGCTGGAAGTTAT 4362
QY      121 TTGTTTGTGAATCCATCTGCTT 144
DB      4361 CCGGTTTGTGTCATGCTTCTT 4338

RESULT 9
V72993/c
ID      V72993 standard; cDNA; 4483 BP.
AC      V72993;
DE      25-FEB-1999 (first entry)
DE      Tobacco basic beta-1,3-glucanase genomic DNA clone pBS-gluc39.1.
DE      Regulation: transcription; plant tissue; chimetic construction; PR:
KW      pathogenesis-related protein; anti-pathogenic; transgenic plant;
KW      beta-1,3-glucanase activity; pest resistance; ss.
OS      Nicotiana sp.
SQ      US5847258-A.

```





```

FT      /tag- h
FT      /note- "coding sequence for B3 gene"
FT      25152. 25157
FT      /tag- 1
FT      /note- "poly-A signal for B3 gene"
FT      prlm_transcript 23748. 31827
FT      /tag- j
FT      /label- B4 gene
FT      29808. 30848
FT      /tag- k
FT      /note- "coding sequence for B4 gene"
FT      31822. 31827
FT      /tag- 1
FT      /note- "poly-A signal for B4 gene"
FT      prlm_transcript 31403. 39691
FT      /tag- m
FT      /label- B1 gene
FT      31403. 31474
FT      /tag- n
FT      /note- "exon 1 of B1 gene"
FT      31475. 33957
FT      /tag- o
FT      /note- "intron 1 of B1 gene"
FT      33958. 34062
FT      /tag- p
FT      /note- "exon 2 of B1 gene"
FT      34063. 35056
FT      /tag- q
FT      /note- "intron 2 of B1 gene"
FT      35057. 35139
FT      /tag- r
FT      /note- "exon 3 of B1 gene"
FT      35140. 38087
FT      /tag- s
FT      /note- "intron 3 of B1 gene"
FT      38088. 39691
FT      /tag- t
FT      /note- "exon 4 of B1 gene"
FT      38148. 39191
FT      /tag- u
FT      /note- "coding sequence for B1 gene"
FT      39674. 39679
FT      /tag- v
FT      /note- "poly-A signal for B1 gene"

PN      WO9746710-A1.
PD      11-DEC-1997.
PD      05-JUN-1997; U09774.
PR      25-APR-1997; US-846111.
PR      05-JUN-1996; US-658578.
PA      (LUDW-) LUDWIG INST CANCER RES.
PI      Boon T, Brasseur F, Lurquin C;
DR      WPI: 98-042216/04.

PT      Screening for the possibility of testicular seminoma and other
PT      cancers - using MAGE-Xp gene primers
PT      Claim 7: Pages 22-34; 40pp; English.
CC      This is the MAGE-B cluster DNA sequence. The invention provides a novel
CC      method for screening for the possibility of a testicular seminoma,
CC      non-small cell lung carcinoma, melanoma, breast cancer, sarcoma, or
CC      leukemia in a sample. The method comprises contacting the sample with
CC      at least one nucleic acid molecule which hybridizes to mRNA corresponding
CC      to an MAGE-Xp gene, and determining hybridization as a determination of
CC      possible presence of testicular seminoma, non-small cell lung carcinoma
CC      melanoma, breast cancer, sarcoma, or leukemia in the sample. By assaying
CC      for the MAGE-B1 gene, or MAGE-B2 gene, both contained within this 40352
CC      base pair MAGE-B cluster DNA sequence, the presence of such cancers in a
CC      sample can be determined. The genomic DNA that encodes the MAGE-B2 gene
CC      consists of nucleotides 3266-7979 of this MAGE-B cluster DNA sequence.
CC      The nucleic acid molecule that encodes a MAGE-B1 gene, consists of, in
CC      5' to 3' order, nucleotides 31403-31474, 33958-34062, 35057-35139, and
CC      38088-39691, nucleotides 31403-31474, 33958-34062, 35057-35139, and
CC      nucleotides 35057-35139 and 38088-39691, and nucleotides 33958-34062 and
CC      38088-39691 of this sequence. The primers specific for the MAGE-B
CC      gene can be used in a kit to amplify a MAGE-B gene.

```

```

SQ      Sequence      40352 BP;      11599 A;      8654 C;      9772 G;      10327 T;

Query Match
Best Local Similarity 8.7%; Score 32; DB 1; Length 40352;
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY      32 CTTTCACTGCAATTTGCAAGGATACATTCCTGATGATATGTTGTCGAA 91
DB      19244 CTTCGCTCTCCCAAAAGTTGGATTCAGAGCGTACAGACCCCTGCTTAAATTA 19185
QY      92 AAAAAAAGTCTTGTGTTAAATTAATTAATTCCTGTTGTCATCTTCTTTCC 151
DB      19184 AAAAAAAGTTTATTTATTTATTTATTTATTTATTTATTTGAGATGAGTCTGCTGCACC 19125

RESULT 15
X15658
ID      X15658 standard; DNA; 1407 BP.
AC      X15658;
DE      07-MAY-1999 (first entry)
DE      Protein phosphatase gene coding sequence of Saccharomyces cerevisiae.
KW      Protein phosphatase gene; growth; fermentation activity;
OS      Saccharomyces cerevisiae.
PN      J11042090-A.
PD      16-FEB-1999.
PE      29-JUL-1997; 203652.
PR      29-JUL-1997; JP-203652.
PA      (KANF ) KANEKA CORP.
PA      (SHOS ) SHOWA SANGYO CO.
DR      WPI: 99-197822/17.

PT      New yeast of controlled activation at low temperatures - useful for
PT      improving the quality of dough
PT      Claim 5: Page 21-22; 41pp; Japanese.
CC      The present sequence represents a protein phosphatase gene sequence
CC      S. cerevisiae in which the growth and/or the fermentation activity
CC      is controlled at least in the range of 0-20 degrees Celsius. These
CC      yeast are prepared by deleting the function of at least one protein
CC      phosphatase gene. The yeast is useful in the production of dough.
SQ      Sequence      1407 BP;      434 A;      222 C;      325 G;      426 T;

Query Match
Best Local Similarity 8.7%; Score 32; DB 1; Length 1407;
Matches 101; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY      21 TGAATTAATTCCTTCAACTGCAATTTGCAAGGATTAACATTCACGTGATATAT 80
DB      891 TGAATTCGACTCAAAATGTTGAGCGTATGAGATCAAAAATTAACAAATTCGAAACGTC 950
QY      81 TGTGTGCAAAAAAAGTCTTTGTTTAAATTAATTCCTGTTGTCATTCATCTT 140
DB      951 TTTTGGCAAGAAGAAAGTATTTTGTATTCATGATTTTTCGAGATGATATACGA 1010
QY      141 GCTTTTCCCATTTGAAGTATGATTAACCAATCTCTGAATGTGAGAAAAACATCTGA 200
DB      1011 AGGTTCGCAATTAACCAACCAAAAAATTAACAGACCCCTGGAATCGTAGTAAGATATGA 1070
QY      201 AGAGCTAGTCTATACCACTCTGACAGTGAATGA 236
DB      1071 CGACATGGAATGATGATCTGTATACCGAATTGA 1106

```

Search completed: September 28, 1999, 12:27:24  
 Job time: 5264 sec







Db 1163 YY 1222  
Oy 122 TGGTTGGTAATCATCTGTTGTTTCCCATCTGATCTGATTAACCATCTGAA 181  
Db 1223 YY 1282  
Oy 182 CTGCTAGAAAAACATCTGAGAGCTAGTCTATCATGACATCTGACAGTGAATGGATGTT 241  
Db 1283 YY 1342  
Oy 242 CTCGAAACATTTACCCAGACAGCTGTTCTATCCTGTTTAATTAATTGTTGGGTT 301  
Db 1343 YY 1402  
Oy 302 CTCACATGATACAAACCTGCTCATCTGATTAAGTCTGATGATGAGT 360  
Db 1403 YYGTACCAATTTCTATCTCTTAACT 1461

RESULT 2  
US-08-475-427-5  
Sequence 5, Application US/08475427  
Patent No. 5859340  
GENERAL INFORMATION:  
APPLICANT: DUBOIS, Michel  
APPLICANT: GRISON, Rene  
APPLICANT: LEGUAY, Jean-Jacques  
APPLICANT: PIGNARD, Annie  
APPLICANT: TOPPAN, Alain  
TITLE OF INVENTION: Recombinant gene coding for a protein  
TITLE OF INVENTION: having endochitinase activity or for a precursor thereof.  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington, D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/475.427  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/842,165  
FILING DATE: 01-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/FR91/00607  
FILING DATE: 21-JUL-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 90 09460  
FILING DATE: 24-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16781/564/BEEL  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ. ID NO.: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3012 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 2384..2462

OTHER INFORMATION: /number- 1  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 2617..2697  
OTHER INFORMATION: /number- 2  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(1942..2383, 2463..2616, 2698..3007)  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: 1942..2166  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: join(2167..2383, 2463..2616, 2698..3007,  
LOCATION: 2698)  
US-08-475-427-5  
Query Match 9.3%; Score 34; DB 4; Length 3012;  
Best Local Similarity 48.9%; Pred. No. 0.74;  
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

RESULT 3  
US-08-181-271A-5/C  
Sequence 5, Application US/08181271A  
Patent No. 5614395  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Umes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/181,271A  
FILING DATE: 13-JAN-94  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA: US 07/425,504  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ. ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-181-271A-5

Query Match 9.2% Score 33.6; DB 1; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 1 GCCTTCTCTCCCTCCTGATTAATTCCTTCACTTGCAATTCGAGGATAC 60  
DB 4481 GCCTTTGCTGCTCTGATAGAGCTCTTCCCTTGCGATGTGATGAATGGA 4422

QY 61 CATTGACGTGATGATATGTTGTCGCAAAAAAGTCTTTTAAATAC 120  
DB 4421 CAAATCAGCGGAGCTGTTGGTAAATGATCATCTTACCTGGAAGTAT 4362  
QY 121 TTGGTTGTGAATCATCTTGCTT 144  
DB 4361 CCGGTTGTGTGCGCATGCTTCTT 4338

RESULT 4  
US-08-449-315-5/C  
Sequence 5, Application US/08449315  
Patent No. 5650505  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
TITLE OF INVENTION: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF INVENTION: DNA SEQUENCES AND USES THEREOF  
CORRESPONDENCE ADDRESSES: 106  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,315  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA: 800  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441

```

1      FLING DATE: 21-DEC-1990
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: US 07/425,504
4      FILING DATE: 20-OCT-1989
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: US 07/848,506
7      FILING DATE: 6-MAR-1992
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07//768,122
10     FILING DATE: 27-SEP-1991
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: US 07/580,431
13     FILING DATE: 7-SEP-1990
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: US 07/368,672
16     FILING DATE: 20-JUN-1989
17     PRIOR APPLICATION DATA:
18     APPLICATION NUMBER: US 07/329,018
19     FILING DATE: 24-MAR-1989
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER: US 08/045,957
22     FILING DATE: 12-APR-1993
23     ATTORNEY/AGENT INFORMATION:
24     NAME: Elmer, James Scott
25     REGISTRATION NUMBER: 36,129
26     REFERENCE/DOCKET NUMBER: S-19825/P1/GC 1727
27     TELECOMMUNICATION INFORMATION:
28     TELEPHONE: (919)541-8614
29     TELEFAX: (919)541-8689
30     INFORMATION FOR SEQ ID NO: 5:
31     SEQUENCE CHARACTERISTICS:
32     LENGTH: 4483 base pairs
33     TYPE: nucleic acid
34     STRANDEDNESS: single
35     TOPOLOGY: linear
36     MOLECULE TYPE: DNA
37     US-08-449-315-5

```

Query Match	9.2%	Score	33.6	DB	2	Length	4483
Best Local Similarity	52.1%	Pred. NO.	1.1				
Matches	75	Conservative	0	Mismatches	69	Indels	0
						Gaps	0

OY	1	GCTCTCTCTCC	CCCCCTCTG	GAATTAATTAAT	CTTCACTG	CACTGGAATTTSCA	AGATTTCA	60	
Db	4481	GCTCTTTTGG	CGCTCTG	GAATAGAG	CTCTTTCC	TGCGCTTTG	GAATGATGAAT	44222	
OY	61	CATTTC	ACTG	AGATTAAT	TGCTGTTC	CAAAAAAAAA	AGATGCTCTTT	TTAAAAATAC	120
Db	4421	CAATTC	ACACG	CACTAGTTT	TGGGGT	ATACATCA	AAATAGATCTCTT	TGCTGGAAGTTAT	43622
OY	121	TTGGTTTGG	ATC	ATCACTT	GGCTT	144			
Db	4361	CCCGTTT	GTGTGTGG	CAATGCTTCTT	4338				

RESULT 5  
US-08-444-803-5/C  
: Sequence 5, Application US/08444803  
: Patent No. 5654416  
: GENERAL INFORMATION:  
: APPLICANT: Rals, John A.  
: APPLICANT: Alexander, Danny C.  
: APPLICANT: Beck, James J.  
: APPLICANT: Duesling, John H.  
: APPLICANT: Friedrich, Leslie B.  
: APPLICANT: Goodman, Robert M.  
: APPLICANT: Hamm, Christian  
: APPLICANT: Meins, Jr., Frederick  
: APPLICANT: Montoya, Alice  
: APPLICANT: Moyet, Mary B.  
: APPLICANT: Neuhaus, Jean-Marc  
: APPLICANT: Payne, George B.

APPLICANT: Spertson, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericca C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,803  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT 1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,119  
REFERENCE/DOCKET NUMBER: S-19825/P1/GGC 1727  
TELECOMMUNICATION INFORMATION:

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/444,803  
FILING DATE: 19-MAY-1995  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/C/GC 1727  
TELECOMMUNICATION INFORMATION:











TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-454-876-5

Query Match 9.2%; Score 33.6; DB 3; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GCCTCTCTCCCTCCCTGATTAATTCCTTCACTTGCAATTTGCAAGATTACA 60  
DB 4481 GCCTCTTGTGCTGCTCTCTGATAGAGCTCTTCTTCCCTTGGAAGTGAATGCA 4422  
QY 61 CATTTCAGTGTGATGATATATTGTTGCAAAAAAAGTCTTTTAAATTAC 120  
DB 4421 CAATTCACAGCAGTGTGTTGGTGAAGATCAAAATAGATCTTACGTGAAGTTAT 4362  
QY 121 TTGCTTGTGAATCATCTTGCTT 144  
DB 4361 CCCGTTGTGTGTCATGCTTCTT 4338

## RESULT 11

US-08-457-364-5/c  
Sequence 5, Application US/08457364  
Patent No. 5847258

## GENERAL INFORMATION:

APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10512  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/457,364  
FILING DATE: 31-May-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-Jan-94  
APPLICATION NUMBER: US 08/093,301

FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-457-364-5

Query Match 9.2%; Score 33.6; DB 4; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GCCTCTCTCCCTCCCTGATTAATTCCTTCACTTGCAATTTGCAAGATTACA 60  
DB 4481 GCCTCTTGTGCTGCTCTCTGATAGAGCTCTTCTTCCCTTGGAAGTGAATGCA 4422  
QY 61 CATTTCAGTGTGATGATATATTGTTGCAAAAAAAGTCTTTTAAATTAC 120  
DB 4421 CAATTCACAGCAGTGTGTTGGTGAAGATCAAAATAGATCTTACGTGAAGTTAT 4362  
QY 121 TTGCTTGTGAATCATCTTGCTT 144  
DB 4361 CCCGTTGTGTGTCATGCTTCTT 4338





APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-455-736-5

Query Match 9.2%; Score 33.6; DB 4; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 1 GCCTTCTCTCCCTCCCTCTGAATTAATCTTCACTTGAATTTGCAAGATTACA 60  
DB 4481 GCCTTCTCTCCCTCCCTCTGAATTAATCTTCTCTGCTTGGAGTATGAAATGG 4422  
QY 61 CATTTCAGTGTATATATGTTGCAAAAAAAGTCTTGTAAATTAAC 120  
DB 4421 CAATTACACGAGCTAGTTTGGGTAGATCAAAATAGATCTTAGCTGGAAGTTAT 4362  
QY 121 TTGGTTGTGAATCATCTTGTCTT 144  
DB 4361 CCCGTTGTGTGTCATGCTTCTT 4338

RESULT 15  
US-08-455-073A-1/C  
Sequence 1, Application US/08455073A  
Patent No. 5876949  
GENERAL INFORMATION:  
APPLICANT: Gideon Dreyfuss  
APPLICANT: Mikiko C. Slomi

APPLICANT: Yan Zhang  
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods  
TITLE OF INVENTION: Of Making And Using The Same  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949r1s  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,073A  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: UPN-2201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4362 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 220..2118  
US-08-455-073A-1

Query Match 8.3%; Score 30.4; DB 4; Length 4362;  
Best Local Similarity 55.8%; Pred. No. 7.8;  
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 90 AAAAAAAGTGTCTTGTAAATTAATCTTGGTGTGAATCCATCTGCTTTCC 149  
DB 3322 AAAAAAAGTGTCTTGTAAATTAATCTTGGTGTGAATCCATCTGCTTTCC 3263  
QY 150 CCATTGGAAGTGTCTTGAATCCATCTGGAAGTGTAGAAAA 193  
DB 3262 CAATCTGTGAATCAATTAAGACATATATTGAAGTTACAA 3219

Search completed: September 28, 1999, 11:33:07  
Job time: 2067 sec



**This Page Blank (uspto)**

OM of: US-09-030-606-115 to: PIR\_60.\* out\_format : pfs

Date: Sep 25, 1999 6:28 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-1998 CompuGen Ltd.

# Command line parameters:

-MODEL=framed+np\_model -DEV=xlp  
-O=/cgn2.1/USPFO-spool/US09030606/runat\_24091999\_171617\_29825/app\_query.fasta.1  
-DB=PIR\_60 -OFT=fastn -SUFFIX=rrp -GAP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -GAP=4.500  
-OGAPEXT=0.050 -XGAP=10.000 -XGAPEXT=0.500 -FGAP=6.000  
-FGAPEXT=7.000 -YGAP=10.000 -YGAPEXT=0.500 -DEL=6.000  
-DELXT=7.000 -START=1 -MATRIX=blomsu62 -TRAN=human40.cdl  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=escore -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09030606 -NCPU=6 -ICPU=3 -WAIT  
-THREADS=1

## Search information block:

Query: US-09-030-606-115  
Query length: 366  
Database: PIR\_60.\*  
Database sequences: 122810  
Database length: 40065486  
Search time (sec): 182.540000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
PIR2:E71422	71.50	136.19	0.6760	1052	hypothetical protein - Arabid
PIR2:PC1123	66.00	127.35	3.40	651	hypothetical protein - bloodfi
PIR2:S75651	66.00	124.17	3.42	972	cation-transporting ATPase (EC
PIR2:S28293	65.00	125.15	4.56	643	hypothetical protein ZC84.3a - C
PIR2:S28294	65.00	124.14	4.57	730	hypothetical protein ZC84.3a - F
PIR2:B71030	63.50	134.02	6.91	136	hypothetical protein Pfl534 - F
PIR2:B46535	63.00	121.98	8.20	537	interleukin 2 receptor beta cha
PIR2:S73727	62.50	118.43	9.56	726	MG307 homolog H08.orf726 - Myc
PIR2:S37837	62.50	115.27	9.62	1082	suppressor protein SP23 - yea
PIR2:SC4057	61.50	120.40	12.71	424	probable C-14 sterol reductase
PIR2:S21166	61.50	119.84	12.73	455	adenylosuccinate synthetase - h
PIR1:YU0U	61.00	125.58	14.53	804	sucrose synthase (EC 2.4.1.13)
PIR2:S57642	61.00	122.46	14.63	283	interferon precursor - duck
PIR2:A44814	61.00	122.43	14.63	284	endoplasmic reticulum protein ftab - Le
PIR1:KART	61.00	107.85	15.11	1785	alpha-casein precursor - rat
PIR2:S53976	60.50	115.53	17.17	586	probable membrane protein YMK3
PIR2:S56969	60.50	114.49	17.21	668	TPP1 protein homolog YJL186w -
PIR2:C24785	60.00	119.13	19.70	322	hypothetical protein 668 - slm
PIR2:Q01191	60.00	117.10	19.79	416	vanH protein - Enterococcus fae
PIR2:S39684	60.00	115.05	19.88	539	ywmn protein - Bacillus subtili
PIR2:S46742	60.00	114.45	19.90	581	interleukin-2 receptor beta cha
PIR2:S59712	59.50	120.69	22.69	229	hypothetical protein YHR072w -
PIR1:D43685	59.50	116.09	22.92	409	nonstructural protein V - parae
PIR2:S55178	59.50	112.82	23.09	617	probable sphingolipid transport
PIR2:S19254	59.50	105.90	23.44	1476	nitrate reductase (NADH) (EC 1.
PIR2:A45773	59.00	128.29	25.80	76	kelch protein, long form - fru
PIR2:S59567	59.00	111.07	26.79	666	histone H2A - rat (fragment)
PIR2:S59591	58.50	112.92	30.85	456	NADH-ubiquinone oxidoreductase
PIR2:A53162	58.50	109.85	31.06	672	adenylosuccinate synthase (EC 6
PIR2:S61157	58.00	111.84	35.75	452	hypothetical protein YR356c -
PIR1:AJMSPS	58.00	109.05	35.97	643	adenylosuccinate synthase (EC 6
PIR1:WML42	57.50	119.93	40.59	1441	EL protein - human papillomavir
PIR2:A48769	57.50	112.70	41.25	351	SLP(w7) alpha-chain - mouse (fr
PIR2:S69218	57.50	111.33	41.37	417	conserved hypothetical protein
PIR2:C69795	57.50	104.21	42.02	1023	pecl12-like protein homolog yed
PIR2:A47296	57.00	121.59	46.76	99	thiazide-sensitive electronentr
PIR2:S28701	57.00	114.11	47.53	254	hypothetical protein 19 - Agrob
PIR2:S40702	57.00	112.93	47.66	295	hypothetical protein CO7A9.10 -
PIR2:S46718	57.00	112.48	47.70	312	hypothetical protein YHR088w -
PIR2:S43181	57.00	112.48	47.70	312	outer membrane protein class 3
PIR2:S45065	57.00	112.48	47.70	312	class 3 outer membrane protein

PIR2:G71128 - 57.00 111.85 47.77 338 | hypothetical protein PH0796  
PIR2:S75651 + 57.00 107.81 48.20 562 | hypothetical protein sll1879  
PIR2:S56681 - 57.00 107.40 48.24 592 | beta-fructofuranosidase (EC

seq\_name: PIR2:E71422

## seq\_documentation\_block:

hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998

C:Accession: E71422  
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D  
P.; Medler, H.; Medler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; G  
avanah, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B  
Nature 391, 485-488, 1998

A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Pulgömen  
erhoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;  
C.; Chalvatzis, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t  
A:Reference number: A71400; MUID:98121113  
A:Accession: E71422  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1052 <BEV>  
A:Cross-references: GB:297339; NID:9244901; PID:e327481; PID:9244929  
C:Genetics:  
A:Map position: 4COP9-4G3845

## alignment\_scores:

Quality: 71.50 Length: 119  
Ratio: 1.254 Gaps: 6  
Percent Similarity: 47.899 Percent Identity: 26.891

## alignment\_block:

US-09-030-606-115/rev x E71422 ..  
Align seg 1/1 to: E71422 from: 1 to: 1052

```

360 ACTTCAGATCAGACGCTTTATGTGACAGATTGGACGAGGTTGTTATG 311
|||||:|||||:|||||
323 ThrlysyshlslsValleu..... 329
310 CATGTAGAGAACCAACTATTTATTAACAGATGAAACAGGCTGT 262
|||||:|||||:|||||
330 .....serthrglyglnleuAlaArgAspAsnSerAlaCysG 342
261 ..CTGGTGAATAGTCTGACAGCATTCACATCACCCTGCAGATGCTG 214
|||||:|||||:|||||
342 LylleuglyglYAlaSerleuthrThraspglypnehr..... 354
213 ATAGACTAGCTCTTCAGATGTTTTCACACAGTTCAGAGATGGTTAATG 164
|||||:|||||:|||||
355 .....ThrlsSerAsnleu..... 359
163 ACTAGTTCATGGGGAAGAACAGATGATTCACAACCAACTAATTT 114
|||||:|||||:|||||
360 .....GlnGlnGlyCysArgLysAspGlyGlnSerAspArgVala 374
113 TAAACAAGACACTTTTTCATGACACATCATATACATCAGAGGAA 64
|||||:|||||:|||||
374 laanhlsllelylspheValGlyleu...HsilaVallylGlyllyls 389
63 ATGTGTAATCCTTCATTCAGATTCAGATTCAGATTCAGATTCAGAGG 14
|||||:|||||:|||||
390 Thrlys...ProthrlyslleuglnGlnAsnArgSerlylArgIleGl 405
|||||:|||||:|||||
405 yAspLys 407
seq_name: PIR2:PC1123

```

seq\_documentation\_block:  
hypothetical protein - bloodfluke planorb (fragment)  
C:Species: Biomphalaria glabrata (bloodfluke planorb)  
C>Date: 17-Apr-1993 #sequence\_revision 17-Apr-1993 #text\_change 17-Mar-1999  
C:Accession: PC1123  
R:Knight, M.; Miller, A.; Raghavan, N.; Richards, C.; Lewis, F.  
Gene 118, 181-187, 1992  
A>Title: Identification of a repetitive element in the snail Biomphalaria glabrata: Rela  
A:Reference number: PC1123; MUID:92280502  
A:Accession: PC1123  
A:Molecule type: DNA  
A:Residues: 1-651 <KN1>  
A:Cross-references: EMBL:X60372

alignment\_scores:  
Quality: 66.00 Length: 96  
Ratio: 1.347 Gaps: 5  
Percent Similarity: 51.042 Percent Identity: 29.167

alignment\_block:  
US-09-030-606-115 x PC1123 ..

Align seg 1/1 to: PC1123 from: 1 to: 651

```
41 TGCATTTGCAGAGATTACACATTTCAGTGTATTTGTTGCAA 90
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
566 CysAsnLeuArgLysLeuHisLeuAsn.....ProGluLeuHeh1 580
91 AAAAAAAGTGTCTTTGTTTAAATTAATTAATTAATTAATTAAT 140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
580 sLysLysLysIleProValValIleThrLysPheLeuGlyLeuThr 597
141 GC.....TTTTCCCATTTGGAATTAATTAATTAATTAATTAAT 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
597 euAspSerLysPheAsnPheLeuProtyr.....Ile 607
176 TCTGAACTGTAGAAAAACATCTGAAGAGCTAGCTATCAGATCGACA 225
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
608 LysGluLeuLysLysGlyGlnLysSerLeuAsnIle.....C 620
226 GGTGAATTGCATGTTCTCAGAACATTTCA.....C 257
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
621 .....LeuArgValLeuSerHisThrAspTyrGlyAla 632
258 CCAGACAGCTGTTTCTATCCCTGTTAATTAATTAATTAATTT 295
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
632 sPaRgAspThrLeuLeuLeuLysTyrSerLeuIle 644
```

seq\_name: p1r2:S77454

seq\_documentation\_block:  
cation-transporting ATPase (EC 3.6.1.-) pacL-2 - Synchocystis sp. (strain PCC 6803)  
N:Alternate names: Protein s111076  
C:Species: Synchocystis sp.  
A:Variety: PCC 6803  
C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 21-Aug-1998  
C:Accession: S77454  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synchocystis  
s.  
A:Reference number: S74322; MUID:97061201  
A:Accession: S77454  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-972 <KAN>  
A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PID:d1018034; PID:g1652379  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C:Genetics:  
A:Gene: pacL-2

C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding doma  
C:Keywords: hydrolase; ion transport  
F:585-757/Domain: ATPase nucleotide-binding domain homology <ATN>

alignment\_scores:  
Quality: 66.00 Length: 53  
Ratio: 2.000 Gaps: 1  
Percent Similarity: 62.264 Percent Identity: 28.302

alignment\_block:  
US-09-030-606-115 x S77454 ..

Align seg 1/1 to: S77454 from: 1 to: 972

```
14 CCGCTCTGATTTAATTTCTTCACTTGCATTTGCAGATTCACAT 63
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
870 ProIleLeuMetAlaIleTyrAlaGlnAlaThrThrMetThrLeuAla 886
64 TTCACGTGATGT.....ATATTGTTGCCAAAAAAG 101
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
886 ILeValAlaCysGlnAspGlyAsnValPheAlaCysArgSerGluArg 903
102 TGTCTTTGTTTAAATTAATTAATTAATTAATTAATTAATTAAT 151
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
903 hrSerAlaPheLysLeuGlyTyrPheSerAsnProLeuIleTyrIleGly 919
152 ATTGGAAT 160
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
920 IleAlaThr 922
```

seq\_name: p1r2:S28293

seq\_documentation\_block:  
hypothetical protein ZC84.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 30-Sep-1993  
C:Accession: S28293  
R:Thomas, K.  
submitted to the EMBL Data Library, December 1992  
A:Reference number: S28285  
A:Accession: S28293  
A:Molecule type: DNA  
A:Residues: 1-643 <THO>  
A:Cross-references: EMBL:Z19157  
C:Genetics:  
A:Introns: 141/2; 167/3; 245/3; 341/2; 412/3; 600/3; 640/3

alignment\_scores:  
Quality: 65.00 Length: 103  
Ratio: 1.226 Gaps: 6  
Percent Similarity: 51.456 Percent Identity: 33.981

alignment\_block:  
US-09-030-606-115/rev x S28293 ..

Align seg 1/1 to: S28293 from: 1 to: 643

```
291 AATTATTAACAGATAGAACAGCGCTGCGTGAATGATGTTCTGAG 243
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 AsnCysLeuAsnArgLeuLeuSerAspAlaSerAsnThrMetValLeu 171
242 .....AACCATCAATTCACCTGTCAGATGCTG 214
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 uIleTyrSerIlePheValAlaArgThrHisSerAspArgLeuSerGluTyr 188
213 ATAGACTAGCTCTTCAGATGTTTCTTACAGATTCAGATGGGTTAAG 164
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 euArgLeuAlaLeuAlaLysIlePhe.....AlaArgLysAlaAlaGlu 202
163 ACTAGTTCATGCGGAAAAACAGATGATTCACAAACAGATTAATT 114
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



```

203 Thr LeuProAsnThrLysGlnIleGlyHisThrLeuAsnValIleL 219
113 TAAACAAGACACTTTTGTTCGACAAACATATACATAC..... 70
219 eu.....Glucysphe.....AsnAlaHisHisGlnLeu 228
69 .....AGTGAATGTGTAATCCTTGCAATTCGACAGTTGAAAG 32
229 ValThrValCysGlnLeuMetCysAspProIleHisLeuMetValProly 245
31 AATTAAA 25
245 sAlaArg 247
seq_name: p1r2:S28294
seq_documentation_block:
hypothetical protein ZC84.3a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 30-Sep-1993
C:Accession: S28294
R:Thomas, K.
Submitted to the EMBL Data Library, December 1992
A:Reference number: S28285
A:Accession: S28294
A:Molecule type: DNA
A:Residues: 1-730 <THO>
A:Cross-references: EMBL:Z19157
C:Genetics:
A:Introns: 77/3; 120/3; 213/1; 228/2; 254/3; 332/3; 428/2; 499/3; 687/3; 727/3

alignment_scores:
Quality: 65.00 Length: 103
Ratio: 1.226 Gaps: 6
Percent Similarity: 51.456 Percent Identity: 33.981

alignment_block:
US-09-030-606-115/rev x S28294 ..
Align seg 1/1 to: S28294 from: 1 to: 730
291 AATTATTAAACAGATAGAAACAGCGCTGTGGTGAATGTTCTAG 243
||| |||||::: ::::: |||||
242 AsnCysLeuAsnArgLeuLeuSerAspAlaSerAsnThrMetValLeuL 258
242 .....AACCATCAATTCACCTGTGAGATCTG 214
258 uIleTyrSerIlePheValArgThrHisSerSerArgLeuSerGluTrpL 275
213 ATAGACTAGCTCTTCAGATGTTTTCACAGATTCAGAGATGGTTAATG 164
||| |||||::: |||||
275 euArgLeuAlaLeuAlaLysLeuPhe.....AlaArgLysAlaAlaGlu 289
163 ACTAGTCCATGGGGAAGAAACAGATGATTCAACAAACCAAGTAATT 114
||| ::::: ||||| ||| ||| ::::: |||||
290 Thr.LeuProAsnThrLysLysGlnIleGlyHisThrLeuAsnValIleL 306
113 TAAACAAGACACTTTTGTTCGACAAACATATACATC..... 70
306 eu.....Glucysphe.....AsnAlaHisHisGlnLeu 315
69 .....AGTGAATGTGTAATCCTTGCAATTCGACAGTTGAAAG 32
316 ValThrValCysGlnLeuMetCysAspProIleHisLeuMetValProly 332
31 AATTAAA 25
332 sAlaArg 334
seq_name: p1r2:D71030
seq_documentation_block:

```

```

hypothetical protein PH1534 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 14-Aug-1998
C:Accession: D71030
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; M01D:98344137
A:Accession: D71030
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-136 <KAM>
A:Cross-references: GB:AP000006; NID:g3236133; PID:d1031587; PID:g3257961
A:Experimental source: strain OT3
A:Note: this accession replaces an Interim accession for a sequence replaced by GenB
C:Genetics:
A:Gene: PH1534

alignment_scores:
Quality: 63.50 Length: 64
Ratio: 1.671 Gaps: 2
Percent Similarity: 59.375 Percent Identity: 31.250

alignment_block:
US-09-030-606-115 x D71030 ..
Align seg 1/1 to: D71030 from: 1 to: 136
147 TCCCATGTGACATGATCAATCAACCATCT.....CTGACTGGTA 187
::: |||||::: |||||::: |||||
25 ThrProValGluThrAlaIleHisProSerPheSerLeuGluHisTrpAl 41
188 GAAACAATCTGAGAGCTAGTCTA.....TCACATCTGACAGTGA 231
::: |||||::: |||||
41 aserAsnIleAspProSerAsnIleArgProSerSerIleAspLysIleI 58
232 TTGGATGCTTCACAGACCATTTCAACCCAGACAGCGCTGTTTCTATCCTGT 281
||| |||||::: |||||
58 LegIlyPheThrProValGlnThrSerProLeuGlyThrThmethIsAla 74
282 TTAATAAATTAGTTGGGTTCTCTACATGATCAACAAACCT 323
75 SerLeuSerMetGlyGlyPheThrIleLeuIleTrnAspPro 88
seq_name: p1r2:B46535
seq_documentation_block:
interleukin 2 receptor beta chain - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 18-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 30-May-1997
C:Accession: B46535
R:Page, T.H.; Dailman, M.J.
Eur. J. Immunol. 21, 2133-2138, 1991
A:Title: Molecular cloning of cDNAs for the rat interleukin 2 receptor alpha and beta
A:Reference number: A46535; M01D:91364784
A:Accession: B46535
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-537 <PAG>
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:60818, NCBIPI:60819)
C:Keywords: cytokine receptor

alignment_scores:
Quality: 63.00 Length: 117
Ratio: 1.212 Gaps: 6
Percent Similarity: 44.444 Percent Identity: 25.641

alignment_block:
US-09-030-606-115/rev x B46535 ..

```

Align seg 1/1 to: B46535 from: 1 to: 537

```

310 CATGTAGAGAACCAACTAATTATTATTAACAGATGAAACAGCGTGTG 261
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 H1s1ePr0sPser0SerGluPhePheSerGln.....LeuSe 295
260 TGGGTGAATGGTCT..... 245
295 rSerGlnH1s1eGlyAspLeuGlnLysTrpLeuSerSerProValProG 312
244 .....GAGAACATCAATTCAC 227
312 InSerPhePheSerProThrGlySerAlaProGluL1eSerProLeuGln 328
226 CTGTCACATGCTGTAGACTAGCTCTTCAGATGTTTCTACACGTTGAG 177
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
329 ValLeuAspArgAspSerLysThrMetGlnMetLeuPheGlnLysG1 345
176 AGATGGGTTAATGACTAGTTCATGGGGAAGCAAGATGGA...TTC 130
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
345 ulsAlSerSerProSerProSerGlyHisSerGlnAlaSerCysPhe 361
129 ACAAAACCAAGTAATTTAAACAAAGACACTTTTCTTTTTCACACAA 80
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 ThrAsnGln.....GlyTyrPhePhePheHisLeuSerAs 373
79 TATACATCACAGTGAATGTCT.....AATCCTTGC 49
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
373 nAlaLeuGlnL1eGlnSerCysGlnValTyrPheTyrTyrAspProCys 389
seq_name: p1r2:573727

```

```

seq documentation_block:
MG307 homolog H08_crf726 - Mycoplasma pneumoniae (ATCC 29342) (SGC3)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #ext_change 17-Jul-1998
C:Accession: S73727
R:Himmelfeich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73727; MUID:97105885
A:Accession: S73727
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-726 <HIN>
A:Cross-references: EMBL:AE000039; GB:U00089; NID:g1674082; PID:g1674084
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Genetic code: SGC3

```

alignment\_scores: Quality: 62.50 Length: 46  
Ratio: 1.953 Gaps: 1  
Percent Similarity: 69.565 Percent Identity: 32.609

alignment\_block: US-09-030-606-115/rev x S73727 ..

Align seg 1/1 to: S73727 from: 1 to: 726

```

268 AGGCTGCTGGGTGAATGTTCTGAGAACCATTCACCTGTCAGA 219
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
635 ArgAspSerGlyLysSerGlyThrGlyAsnThrLeuThrThrAspG1 651
218 TGGTATAGACTAGCTCTTCAGATGTTTCTTACAGATTCAGAGATGGGT 169
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
651 ySerAspLysLeuThrTyrLeuGlnPheSerTyrLysAspL1eAspGlyL 668
168 TA..ATGACTAGTTCATGGGGAAGCAAGATGCG 134
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

668 euSerLeuSerSerSerAsnGlyThrSerSerLysPhe 680

```

seq_name: p1r2:S37837
seq documentation_block:
Suppressor protein SPT23 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YKL020C
C:Species: Saccharomyces cerevisiae
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 10-Jul-1998
C:Accession: S37837; S40696
R:Rieger, M.
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37832
A:Accession: S37837
A:Molecule type: DNA
A:Residues: 1-1082 <RIE>
A:Cross-references: EMBL:D28020; NID:g4486010; PID:g4486011; MIPS:YKL020C
A:Experimental source: strain S28C
R:Burkett, T.J.; Garfinkel, D.J.
Yeast 10, 81-92, 1994
A:Title: Molecular characterization of the SPT23 gene: a dosage-dependent suppressor
A:Reference number: S40695; MUID:94262317
A:Accession: S40696
A:Molecule type: DNA
A:Residues: 136-714, 'P', 716-737, 'ERKIVLVLTATYGVY' <BUR>
A:Cross-references: EMBL:L24760; NID:g404815; PID:g404817
C:Genetics:
A:Gene: SGD:SPT23
A:Cross-references: SGD:S0001503; MIPS:YKL020C
A:Map position: 11L
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology
F:742-774/Domain: ankyrin repeat homology <ANI>

```

alignment\_scores: Quality: 62.50 Length: 104  
Ratio: 1.157 Gaps: 7  
Percent Similarity: 51.923 Percent Identity: 30.769

alignment\_block: US-09-030-606-115/rev x S37837 ..

Align seg 1/1 to: S37837 from: 1 to: 1082

```

268 AGGCTGCTGGGTGAATGTTCTGAGAACCATTCACCTGTCAGA 219
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
141 ArgLeuArgAsnAlaHisAlaArgLysAsnPhePheHisLeu...Pr 156
218 TGGTATAGACTAGCTCTTCAGATGTTTCTTACAGATTCAGAGATGGGT 169
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
156 oSerAspCysIleAlaLysAspLysPhePhe.ThrSerSer..... 169
168 TAATGACTAGTTCATGGGGAAGCAAGATGATTCACAAACCAAGT 119
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
170 .....AspAspProThrAsnLeuTh 176
118 AATTTTAACAAGAC.....ACTTTT.....TTT 90
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
176 r1leProAsnArgAspL1eAsnGlnLArgThrLeuPheLeuAspAlaPhe1 193
89 TG.....CAACACAAATATACATCAGTGAATG 61
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
193 euLeuCysAlaSerAsnAsnSerAsnAsnPhelysGlnTrpVal 209
60 TGTATCTCTTGAATTTGCAAGTGAAGAATTTAAATTCAGAGAGGGA 11
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
210 CysAsnArgCys.....IleAsnArgGlnLysArgArgLaseSer 223
10 GAGAAAGAC 1
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
223 gArgLysSer 226
seq_name: p1r2:JC4057

```



US-09-030-606-115/rev x A44814



**This Page Blank (uspto)**

OM of: US-09-030-606-115 to: SwissProt\_37:\* out\_format: pfs  
Date: Sep 25, 1999 4:45 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODEL-frame+ n2p.model -DEV-xip  
-O-cgnt2.1/USPTO.SP00L/US09030606/runatc\_24091999.171618.29883/app-query.fasta.1  
-DS-SwissProt\_37 -QFMT-fastan -SUFFIX-rsp -GAPOP-12.000  
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000  
-OGAPOP-6.500 -OGAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500  
-FGAPOP-6.000 -FGAPEXT-7.000 -YGAPO-10.000 -YGAPEXT-0.500  
-DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-blosum62  
-TRANS-human0.cdi -LIST-45 -DOCALLIGN-200 -THR SCORE-esscore  
-ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-stat -USER-US09030606  
-NCPU-6 -ICPU-3 -WAIT -THREADS-1

Search information block:  
Query: US-09-030-606-115  
Query length: 366  
Database: SwissProt\_37:\*\br/>Database sequences: 77977  
Database length: 28268293  
Search time (sec): 155.430000

```
score_list:
Sequence      Strd Orig  Zscore  EScore Len  Documentation
SwissProt_37:IAP.CHICK + 65.00 125.17 3.38 611 090660 caenorhabditis elegans
SwissProt_37:YH83.CAEEL - 65.00 123.67 3.43 730 003609 caenorhabditis elegans
SwissProt_37:YH2B.RAT - 63.00 121.69 6.00 537 026896 rattus norvegicus (rat)
SwissProt_37:SR23.YEAST - 62.50 114.65 7.35 1082 035210 saccharomyces cerevisiae
SwissProt_37:YH2B.CAEEL - 61.50 119.67 9.19 455 030520 homo sapiens (human)
SwissProt_37:YH2B.CAEEL - 61.00 114.87 9.61 804 000917 arabidopsis thaliana
SwissProt_37:YH2B.CAEEL - 61.00 125.84 9.92 191 051526 anas platyrhynchos (dc)
SwissProt_37:YH2B.CAEEL - 61.00 122.50 10.24 284 026815 rattus norvegicus (rat)
SwissProt_37:YH2B.CAEEL - 61.50 102.08 10.85 3672 021313 caenorhabditis elegans
SwissProt_37:YH2B.CAEEL - 60.50 129.12 11.01 113 048358 drosophila melanogaster
SwissProt_37:YH2B.CAEEL - 60.50 107.02 11.86 1785 004952 saccharomyces cerevisiae
SwissProt_37:YH2B.CAEEL - 60.50 115.26 12.56 586 004652 saccharomyces cerevisiae
SwissProt_37:YH2B.CAEEL - 60.00 119.16 13.86 332 005709 enterococcus faecium
SwissProt_37:YH2B.CAEEL - 60.00 117.00 14.14 416 039597 bacillus subtilis (hy)
SwissProt_37:YH2B.CAEEL - 60.00 114.82 14.44 539 016287 mus musculus (mouse)
SwissProt_37:YH2B.CAEEL - 60.00 114.19 14.53 581 038767 saccharomyces cerevisiae
SwissProt_37:YH2B.CAEEL - 59.50 120.89 15.61 229 021740 human parainfluenza 4b
SwissProt_37:YH2B.CAEEL - 59.50 117.99 16.04 333 018671 bos taurus (bovine)
SwissProt_37:YH2B.CAEEL - 59.50 116.00 16.35 409 047013 saccharomyces cerevisiae
SwissProt_37:YH2B.CAEEL - 58.50 111.61 17.04 689 004652 drosophila melanogaster
SwissProt_37:YH2B.CAEEL - 58.50 113.22 22.01 434 010023 caenorhabditis elegans
SwissProt_37:YH2B.CAEEL - 58.50 112.81 22.09 456 046664 mus musculus (mouse)
SwissProt_37:YH2B.CAEEL - 58.00 113.31 25.17 375 063447 rattus norvegicus (rat)
SwissProt_37:YH2B.CAEEL - 58.00 111.65 25.58 457 028650 mus musculus (mouse)
SwissProt_37:YH2B.CAEEL - 58.00 108.77 26.28 643 027221 human papillomavirus 1
SwissProt_37:YH2B.CAEEL - 57.50 107.38 30.50 663 055439 bovine adenovirus type 1
SwissProt_37:YH2B.CAEEL - 57.50 103.72 31.57 1023 055019 pseudopneumonia typh
SwissProt_37:YH2B.CAEEL - 57.00 114.31 32.69 254 034321 caenorhabditis elegans
SwissProt_37:YH2B.CAEEL - 57.00 113.05 33.09 295 038805 saccharomyces cerevisiae
SwissProt_37:YH2B.CAEEL - 57.00 108.88 34.42 484 087187 candida maltosa (yeast)
SwissProt_37:YH2B.CAEEL - 57.00 108.80 34.45 489 028793 pototous tridactylus
SwissProt_37:YH2B.CAEEL - 57.00 107.19 34.98 582 039689 daucus carota (carrot)
SwissProt_37:YH2B.CAEEL - 56.50 115.44 37.03 194 030080 hamster polyomavirus
SwissProt_37:YH2B.CAEEL - 56.50 109.89 39.04 375 061041 mus musculus (mouse)
SwissProt_37:YH2B.CAEEL - 57.00 95.62 39.04 2339 027655 plasmodium falciparum
SwissProt_37:YH2B.CAEEL - 56.50 109.33 39.25 401 030799 hamster polyomavirus
SwissProt_37:YH2B.CAEEL - 56.50 109.31 39.26 502 050529 avian sarcoma virus (c)
SwissProt_37:YH2B.CAEEL - 56.50 107.35 39.99 407 008942 gallus gallus (chicken)
SwissProt_37:YH2B.CAEEL - 56.00 120.68 40.35 91 078424 galliardia theta (chicken)
SwissProt_37:YH2B.CAEEL - 56.50 106.37 40.37 570 034520 caenorhabditis elegans
SwissProt_37:YH2B.CAEEL - 56.50 106.34 40.38 572 041199 homo sapiens (human)
SwissProt_37:YH2B.CAEEL - 56.50 106.34 40.38 572 037427 mus musculus (mouse)
```

SwissProt\_37:DEY1.RAT - 56.50 106.34 40.38 572 062950 rattus norvegicus (r)  
SwissProt\_37:CTRO.MOUSE - 56.50 97.69 43.83 1597 049025 mus musculus (mou)  
SwissProt\_37:YAG1.YEAST + 56.00 107.86 45.57 417 039713 saccharomyces cere

seq\_name: SwissProt\_37:IAP.CHICK

seq\_documentation\_block:  
ID IAP.CHICK STANDARD; PRT; 611 AA.

AC 090660;  
DT 01-NOV-1997 (REL. 35, CREATED)  
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)  
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
DE INHIBITOR OF APOPTOSIS PROTEIN (IAP) (INHIBITOR OF T CELL APOPTOSIS  
DE PROTEIN).  
GN ITA.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;  
OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE-SPLEEN;  
RX MEDLINE, 97101112.  
RA DIBBY M.R., KIMPTON W.G., YORK J.J., CONNICK T.E., LOWENTHAL J.W.;  
RT "ITA, a vertebrate homologue of IAP that is expressed in T  
RT lymphocytes."  
RL DNA CELL BIOL. 15:981-988(1996).  
CC -1- FUNCTION: APOPTOTIC SUPPRESSOR (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY NUCLEAR.  
CC -1- TISSUE SPECIFICITY: CELLS OF THE T LYMPHOCYTE LINEAGE. FOUND IN  
CC BOTH CORTICAL AND MEDULLARY CELLS OF THE THYMUS.  
CC -1- DEVELOPMENTAL STAGE: HIGH LEVELS ARE INDUCED WITHIN 4-8 HOURS OF  
CC T-CELL ACTIVATION IN SPLEEN AND THYMUS.  
CC -1- SIMILARITY: MEMBER OF THE IAP FAMILY.  
CC -1- SIMILARITY: CONTAINS 3 BIR DOMAINS (BACULOVIRAL INHIBITION OF  
CC APOPTOSIS PROTEIN REPEAT).  
CC -1- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----

DR EMBL, U27466; G1335774;  
DR PROSITE; PS01282; BIR\_REPEAT; 3.  
DR PFM; PFM00653; BIR; 3.  
KW APOPTOSIS; ZINC-FINGER; REPEAT; NUCLEAR PROTEIN.  
FT REPEAT 30 97 BIR\_REPEAT 1.  
FT REPEAT 176 242 BIR\_REPEAT 2.  
FT REPEAT 262 329 BIR\_REPEAT 3.  
FT ZN\_FING 564 598 C3HC4-TYPE.  
SQ SEQUENCE 611 AA; 69009 MW; 4E03CCFC CRC32;

alignment\_scores:  
Quality: 65.00 Length: 176  
Ratio: 1.048 Gaps: 7  
Percent Similarity: 35.227 Percent Identity: 24.432

alignment\_block:

US-09-030-606-115 x IAP.CHICK ..

Align seg 1/1 to: IAP.CHICK from: 1 to: 611  
9 TCTCCCTCTGATTAATTAATCTTCACTGCAATTCAGAGAT... 56  
|||||  
7 SerPDPDeuA[aserVal]MetLysGlnAsnAlaHisCysGlyIcdu 23  
|||||  
57 ....TACACATTCCTGATGATGATATAT..... 80  
|||||

```

23 ulystYrAspPheSerCysGluLeuYrIrIgmetsertInPheSerThp 40
80 .....
40 heProValAsnValProValSerGluArgArgLeuAlaArgAlaGlyPhe 56
81 ..... TGTGTGCAAAAAAAAAGTGT..... 104
57 TyrTyrThrGlyValGlnAspLysValLysCysPheSerCysGlyLeuVa 73
104 ..... 104
73 lleuAspAsnTrpGlnProGlyAspAsnAlaMetGluLysHisLysGlnV 90
105 ..... CTTGTGTTAAATTACTT 122
90 aLYrProSerCysSerPheValGlnAsnMetLeuSerLeuAsnAsnLeu 106
123 GGTTCATGATCCATCTGCTTTCCATCCATGGAAGTGTGTC... ATTA 169
107 GlyLeuSerThrHisSerAlaPheSerProLeuValAlaSerAsnLeu 123
170 CCCATCTGCAATCGGTAGAAAAACATCTGAAGAGCTAGTCTACAGAT 219
123 rProSerLeuArg..... SerMetThrLeuSerProS 134
220 CTGACAGCGTGATGGATGGATGCTCTCAGAACCATTTCA..... CCCAGA 262
134 eArPheGluGlnValGlyTyrPheSerGlySerPheSerSerPheProArg 150
263 CAGCCGCTGTCTATCCTGTTTAAATTAAGTTGGGTTCTCTACATGCA 312
151 AspProValThr..... 154
313 TAACAACCGCTCCATCTGTACAT 340
155 .ThrArgAlaAlaGluAspLeuSerHis 163

seq_name: Swissprot_37:YN83_CAEEL

seq_documentation_block:
ID YN83_CAEEL STANDARD: PRt: 730 AA.
AC 003609:
DT 01-FEB-1994 (REL. 28, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE HYPOTHEtical 80.7 KD PROTEIN ZC84.3 IN CHROMOSOME III.
CN ZC84.3.
OS CAENORABDITIS ELEGANS.
OC EUKARIOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2:
RX MEDLINE: 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BEKKS M.,
RA BONEFIELD J., BURTON J., CONNELL M., COPEY T., COOPERSON A.,
RA CRAXTON M., DEAR S., DU Z., DUBREIN R., FAVELLO A., FRASER A.,
RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLER L., JIER M.,
RA JOHNSON L., JONES M., KERSHAW J., KIRSTEN J., LAISSNER N.,
RA LARREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
RA PAXSONS J., PERCY C., REIKEN L., ROOPRA A., SAUNDERS E., SHOWNKEEN R.,
RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER D., STADEN R.,
RA SUSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA WOHLDMAN P.;
RA "2.2 kb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL NATURE 368:32-38(1994).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on these

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
CC -----
DR EMBL: Z19157; G6930; -
DR PIR: S28294; S28294.
DR PIR: S28293; S28293.
DR WORMPEP: ZC84.3; CE15022.
DR HYPOTHETICAL PROTEIN.
SQ SEQUENCE 730 AA; 80739 MW; 787AD62E CRC32;

alignment_scores:
      Quality: 65.00      Length: 103
      Ratio: 1.226      Gaps: 6
      Percent Similarity: 51.456      Percent Identity: 33.981

alignment_block:
US-09-030-606-115/rev x YN83_CABEL ..

Align seg 1/1 to: YN83_CABEL from: 1 to: 730

291 AATTATTAACAGGATAGAAACAGCGCTGTGGTGAATGTTCTGAG 243
    || |||||::: :: :::: |||||
242 AsnCysLeuAsnArgLeuLeuSerAspAlaSerAsnThrMetValLeuG1 258
242 ..... ACCATCCAAATGCACCTGCAGATGCTG 214
    ::||| ::|||:::
258 uIeYrSerIlePheValArgThrHisSerSerArgLeuSerG1uTrpL 275
213 ATRGACTACTCTTCAGATGTTTCTTACCAGTTCAGAGATGGGTTAATG 164
    |||||::: ||| :::
275 euArgLeuAlaLeuAlaLysLeuPhe.... AlaArgLysAlaIaGlu 289
163 ACTAGTTCGAATGGGGAAAGCAAGATGATTCACAAACCAAGTAATT 114
    || ::||| ||||| || ||| ::|||
290 Thr.LeuProAsnThrLysLysGlnIleGlyHisThrLeuAsnValIleL 306
113 TAAACAAGACACTTTTCTTTTTCGACACCAATATACATCAC..... 70
    || :::: ||| ||| |||||
306 eu.....GlucysPhe..... AsnAlaHisHisGlnLeu 315
69 ..... AGTGAATGTGTAAATCTCTTCCAAATTCGCAAGTTGAAG 32
    :: |||||::: ||| :::
316 ValThrValCysGluLeuMetCysAspProIleHisLeuMetValProL 332
31 AATTAA 25
    : ::
332 SALAArg 334

seq_name: SwissProt_37:IL2B_RAT

seq_documentation_block:
ID IL2B_RAT STANDARD: PRT: 537 AA.
AC P26896;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DE 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE INNERLEUKIN-2 RECEPTOR BETA CHAIN PRECURSOR (IL-2 RECEPTOR) (P70-75)
DE (HIGH AFFINITY IL-2 RECEPTOR BETA SUBUNIT) (CD122).
GN IL2RB.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ROENTHIA; SCUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 91364784.
RA PAGE T.H., DALLMAN M.J.;
RT "Molecular cloning of cDNAs for the rat interleukin 2 receptor alpha
RT and beta chain genes: differentially regulated gene activity in
RT response to mitogenic stimulation.";
RL EUR. J. IMMUNOL. 21:2133-2138(1991).

```



```

seq_name: SwissProt_37:SP23_YEAST
seq_documentation_block:
ID      SP23_YEAST      STANDARD;      PRT;      1082 AA.
AC      P35210;
DT      01-FEB-1994 (REL. 28, CREATED)
DT      01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT      01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE      SP23 PROTEIN.
GN      SP23 OR YK1020C.
OS      SACCCHAROMYCES CREVISIAE (BAKER'S YEAST).
OC      EUKARYOTA: FUNGI: ASCOMYCOTA: HEMIASCOCETES: SACCCHAROMYCETALES;
NC      SACCCHAROMYCETACEAE; SACCCHAROMYCES.
[1]
RP      RIEGER M.;
RL      SUBMITTED (MAR-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RN      SEQUENCE OF 136-752 FROM N.A.
RX      MEDLINE: 94262317.
RA      BURKETT T.J., GARFINKEL D.J.;
RT      "Molecular characterization of the SP23 gene: a dosage-dependent
RT      suppressor of Ty-induced promoter mutations from Saccharomyces
RT      cerevisiae."
RL      YEAST 10:81-92(1994).
CC      -1- FUNCTION: DOSAGE-DEPENDENT SUPPRESSOR OF TY-INDUCED PROMOTER
CC      MUTATIONS. MAY EXERT IT SUPPRESSION EFFECT THROUGH PROTEIN-PROTEIN
CC      INTERACTIONS SINCE DOES NOT PRESENT ANY OF THE MOTIFS GENERALLY
CC      FOUND IN TRANSCRIPTIONAL ACTIVATORS OR DNA BINDING PROTEINS.
CC      -1- SIMILARITY: TO YEAST Y1R033W.
CC      -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 738
CC      ONWARD AND IS SHORTER (752 AA) DUE TO A FRAMESHIFT.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb.slb.ch).
CC      -----
DR      EMBL; Z28020; G486011; -.
DR      EMBL; L24760; G404817; ALT_FRAME.
DR      PIR; S37837; S37837.
DR      SGD; L0002041; SP23.
DR      PFM; PF00023; ank; 2.
DR      CONFLICT 715 715 H -> P (IN REF. 2).
SQ      SEQUENCE 1082 AA; 121337 MW; F69D5306 CRC32;

alignment_scores:
      Quality:      62.50      Length:      104
      Ratio:      1.157      Gaps:      7
Percent Similarity: 51.923      Percent Identity: 30.769

alignment_block:
US-09-030-606-115/rev x SP23_YEAST ..

Align seg 1/1 to: SP23_YEAST from: 1 to: 1082

268 AGGCTGTCTGGGTGAATGTTCTGAGAACATTCACATTCACCTGTGACA 219
      |||:::      :::::      ::|||      |||||
141 AGTLEAGASnAlaHisAlaArgLysAsnPhenIleLeu...Pr 156
218 TCGTGTATAGACTAGCTCTTCAGATGTTTTCACACAGTTCACAGATGGT 169
      ::|||      ::|||      ::|||      |||||
156 oSeraspCysIleAlaLysAspLysPheHe.ThrSerSer..... 169
168 TATATGACTAGTTCAAATGGGGAAGAAAGCAAGATGATTCACAAACCACT 119
      |||      |||||      ::

```

```

170 ..... AspaspProthrasLeuth 176
118 AATTAAACAAGAC.....ACTTTT.....TTT 90
   ||| ||| ||| ||| ||| ||| ||| ||| |||
176 rllrProAsnArgspLrnsncluargrtlrLeuPheLeuAspAlaPheL 193
   ||| ||| ||| ||| ||| ||| ||| ||| |||
89 TG.....CAACACATATACATCAGTGAATG 61
   ||| ||| ||| ||| ||| ||| ||| ||| |||
193 eueucCysAlaSerAsnAsnAsnSerAsnSphelyscInthrTyVal 209
   ||| ||| ||| ||| ||| ||| ||| ||| |||
60 TCCTAATCCCTGCAATTCAGTGAAGATTAATTCAGAGCGCA 11
   ||| ||| ||| ||| ||| ||| ||| ||| |||
210 CysAsnArgYys.....lleAsnArgLulysArgArgAlaSerAr 223
   ||| ||| ||| ||| ||| ||| ||| ||| |||
10 GAGAAAGAC 1
   ||| ||| ||| ||| ||| ||| ||| ||| |||
223 garGlySer 226

seq_name: Swissprot_37:PURA_HUMAN

seq_documentation_block:
ID PURA_HUMAN STANDARD: PRT: 455 AA.
AC P30520:
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ADENYLOSUCCINATE SYNTHETASE (EC 6.3.4.4) (IMP--ASPARTATE LIGASE).
GN ADSS.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER:
RX MEDLINE: 92275078.
RA POWELL S.M., ZALKIN H., DIXON J.E.;
RT "Cloning and characterization of the cDNA encoding human
   adenylosuccinate synthetase.";
RL FEBS LETT. 303:4-10(1992).
RN [2]
RP REVISIONS.
RC TISSUE-LIVER:
RA STONE R.L.;
RL SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE DE NOVO PATHWAY OF PURINE
   NUCLEOTIDE BIOSYNTHESIS.
CC -1- CATALYTIC ACTIVITY: GTP + IMP + L-ASPARTATE -> GDP +
   ORTHOPHOSPHATE + ADENYLOSUCCINATE.
CC -1- PATHWAY: FIRST COMMITTED STEP IN AMP BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, AND PARTIALLY ASSOCIATED WITH
   PARTICULATE FRACTIONS.
CC -1- SIMILARITY: WITH OTHER ADENYLOSUCCINATE SYNTHETASES.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC .....
CC EMBL: X6503; G415849; -.
CC PIR: S21166; S21166.
CC MIM: 103060; -.
CC PROSITE: PS00513; ADENYLOSUCIN_SYN_2; 1.
CC PROSITE: PS01266; ADENYLOSUCIN_SYN_1; 1.
CC PFM: PF00709; Adenylsuc_syn; 1.
CC HSSP: P12283; IADI.
CC PURINE BIOSYNTHESIS; LIGASE; GTP-BINDING.
CC NP_BIND 38 44 GTP (POTENTIAL).
CC ACT_SITE 172 172 BY SIMILARITY.
CC ACT_SITE 179 179 BY SIMILARITY.
CC SEQUENCE 455 AA; 49915 MW; 28632370 CRC32;

```

```

alignment_scores:
  Quality: 61.50      Length: 64
  Ratio: 1.538
  Percent Similarity: 62.500      Percent Identity: 31.250

Alignment_block:
US-09-030-606-115/rev x PURA_HUMAN ..

Align seg 1/1 to: PURA_HUMAN from: 1 to: 455

331 ATTGAGACGGGTTTGTATGCATGTAGAACCAACCAATTTATTAA 282
   ||||| ||| ||| ||| ||| ||| ||| ||| |||
94 lIeGlyAsnclYvalValleHsleu.....ProGlyLeuPheGlul 108
   ||| ||| ||| ||| ||| ||| ||| ||| |||
281 ACAGCATGAAACAGGCTGTCTGGGTGAATGTCTTGAGAACCATCCA 232
   ||| ||| ||| ||| ||| ||| ||| ||| |||
108 uAlaGlyAsnValGlnLysGly...LysGlyLeuGlnGlyTyrGluL 124
   ||| ||| ||| ||| ||| ||| ||| ||| |||
231 TTCACCTGCAGATGCTGATAGCTAGCTCTCAGATGTTTTCACGAG 182
   ||| ||| ||| ||| ||| ||| ||| ||| |||
124 yArGlyLeuIleIleSerAspArgAlaHisIleValPheAspPheHisGln 140
   ||| ||| ||| ||| ||| ||| ||| ||| |||
181 TTCAGAGATGGTTATGACTAGTTCATGCGGGAAGCA 140
   ||||| ||| ||| ||| ||| ||| ||| ||| |||
141 AlAlaAspGlylIleGlnGlnGlnGlnArgGlnGlnGlnAla 154

seq_name: Swissprot_37:SUS2_ARATH

seq_documentation_block:
ID SUS2_ARATH STANDARD: PRT: 804 AA.
AC Q00917:
DT 01-APR-1993 (REL. 25, CREATED)
DT 01-APR-1993 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE SUCROSE SYNTHASE (EC 2.4.1.13) (SUCROSE-UDP GLUCOSYLTRANSFERASE).
DE ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
OC EUKARYOTA; VIRIDIPHYTES; STREPTOPHYTES; EMBRYOPHYTES; TRACHEOPHYTES;
OC EUPHYLOPHYTES; SPERMATOPHYTES; MAGNOLIOPHYTES; EUDICOTYLEDONS; ROSIDAE;
OC CARPALES; BRASSICACEAE; ARABIDOPSIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA:
RX MEDLINE: 92119221.
RA CHOPEA S., DEL-FAVERO J., DOLEFERUS R., JACOBS M.;
RT "Sucrose synthase of Arabidopsis: genomic cloning and sequence
   characterization.";
RL PLANT MOL. BIOL. 18:131-134(1992).
CC -1- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND
   FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.
CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE -> UDP + SUCROSE.
CC -1- INDUCTION: BY ANAEROBIC STRESS.
CC -1- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL Outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (see http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC .....
CC EMBL: X60987; G16526; -.
CC PIR: S19125; YDMU.
CC PFM: PF00534; Glycos_transf_1; 1.
CC PFM: PF00862; Sucrose_synth; 1.
CC TRANSFERASE; GLYCOSYLTRANSFERASE; MULTIGENE FAMILY.
CC SEQUENCE 804 AA; 91989 MW; 9B980B3B CRC32;

alignment_scores:
  Quality: 61.50      Length: 73

```

CC modified and this statement is not removed. Usage by and for commercial entities remains a violation of the license.

```
CC or send an email to license@lsb-sib.ch).
CC -----
CC DR EMBL: J00710: -; NOT_ANNOTATED_CDS.
CC DR EMBL: X03585: G736286: -
CC DR EMBL: X03586: E10128: -
CC DR EMBL: X03587: E10129: -
CC DR EMBL: X03588: E10130: -
CC DR PIR: A03105: KART.
CC DR PROSITE: PS00306: CASEIN_ALPHA_BETA; 1.
CC DR PFAM: PF00363: caseins; 1.
CC KW MILK: PHOSPHORYLATION: REPEAT; SIGNAL.
CC FT SIGNAL 1 15
CC FT CHAIN 16 284 ALPHA CASEIN.
CC FT MOD_RES 34 34 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 35 35 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 36 36 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 92 92 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 93 93 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 94 94 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 95 95 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 96 96 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 97 97 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 98 98 PHOSPHORYLATION (POTENTIAL).
CC FT MOD_RES 99 99 PHOSPHORYLATION (POTENTIAL).
CC FT DOMAIN 138 197 10 x 6 AA TANDEM REPEATS.
CC FT REPEAT 138 143 1.
CC FT REPEAT 144 149 2.
CC FT REPEAT 150 155 3.
CC FT REPEAT 156 161 4.
CC FT REPEAT 162 167 5.
CC FT REPEAT 168 173 6.
CC FT REPEAT 174 179 7.
CC FT REPEAT 180 185 8.
CC FT REPEAT 186 191 9.
CC FT REPEAT 192 197 10.
CC SO SEQUENCE 284 AA: 31801 MW: 57147 FB CRC32;

alignment_scores:
    Quality: 61.00      Length: 49
    Ratio: 1.743      Gaps: 0
    Percent Similarity: 71.429      Percent Identity: 32.653

alignment_block:
US-09-030-606-115/rev x CASI_RAT ..

Align seg 1/1 to: CASI_RAT from: 1 to: 284

189 TCTACAGTTCAGAGATGGGTTAATGACTAGTTCACATGGGGAAGCA 140
    |||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
96 SerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 112
139 AGATGATTCACAAACCAAGTATTTTAAACAAGACACTTTTTTTT 90
    |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
112 nlysaSnllleAlaSnlysgllleLnuSnarGcystHleuGlucInL 129
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
129 eucInArgeInlelystySerGlnleuInGlnAlaSerleu 144

seq_name: Swissprot_37: LML2_CAEEL

seq_documentation_block:
ID LML2_CAEEL STANDARD: PRT: 3672 AA.
AC Q21313:
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE LAMININ-LIKE PROTEIN K08C7.3 PRECURSOR.
GN K08C7.3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL NZ;
RA BERKS M.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 21.5 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 LAMININ DOMAIN IV.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC DR EMBL: Z70286: E1348035: -
CC DR WORMPEP: K08C7.3; CE06136.
CC DR PROSITE: PS00022; EGF_1; 19.
CC DR PROSITE: PS01186; EGF_2; 4.
CC DR PROSITE: PS01248; LAMININ_TYPE_EGF; 21.
CC DR PFAM: PF00052; laminin_B; 1.
CC DR PFAM: PF00053; laminin_EGF; 21.
CC DR PFAM: PF00054; laminin_G; 4.
CC DR PFAM: PF00055; laminin_Nterm; 1.
CC DR HSSP: P02468; IKLO.
CC KW HYPOTHETICAL PROTEIN: LAMININ EGF-LIKE DOMAIN; SIGNAL; REPEAT.
CC FT SIGNAL 1 27
CC FT CHAIN 28 3672
CC FT MOD_RES 28 297 LAMININ-LIKE PROTEIN K08C7.3.
CC FT MOD_RES 298 297 LAMININ N-TERMINAL (DOMAIN VI).
CC FT MOD_RES 357 426 LAMININ EGF-LIKE 1.
CC FT MOD_RES 427 471 LAMININ EGF-LIKE 2.
CC FT MOD_RES 472 518 LAMININ EGF-LIKE 3.
CC FT MOD_RES 519 563 LAMININ EGF-LIKE 4.
CC FT MOD_RES 564 609 LAMININ EGF-LIKE 5.
CC FT MOD_RES 610 655 LAMININ EGF-LIKE 6.
CC FT MOD_RES 656 700 LAMININ EGF-LIKE 7.
CC FT MOD_RES 701 755 LAMININ EGF-LIKE 8.
CC FT MOD_RES 756 808 LAMININ EGF-LIKE 9.
CC FT MOD_RES 809 839 LAMININ EGF-LIKE 10.
CC FT MOD_RES 1415 1460 LAMININ EGF-LIKE 11 (INCOMPLETE).
CC FT MOD_RES 1461 1505 LAMININ EGF-LIKE 12.
CC FT MOD_RES 1506 1553 LAMININ EGF-LIKE 13.
CC FT MOD_RES 1554 1604 LAMININ EGF-LIKE 14.
CC FT MOD_RES 1605 1614 LAMININ EGF-LIKE 15.
CC FT MOD_RES 1615 1796 LAMININ EGF-LIKE 16 (N-TERMINAL).
CC FT MOD_RES 1797 1829 LAMININ DOMAIN IV.
CC FT MOD_RES 1830 1879 LAMININ EGF-LIKE 16 (C-TERMINAL).
CC FT MOD_RES 1880 1936 LAMININ EGF-LIKE 17.
CC FT MOD_RES 1937 1989 LAMININ EGF-LIKE 18.
CC FT MOD_RES 1990 2036 LAMININ EGF-LIKE 19.
CC FT MOD_RES 2037 2083 LAMININ EGF-LIKE 20.
CC FT MOD_RES 2084 2131 LAMININ EGF-LIKE 21.
CC FT MOD_RES 2132 307 LAMININ EGF-LIKE 22.
CC FT MOD_RES 308 320 BY SIMILARITY.
CC FT MOD_RES 321 331 BY SIMILARITY.
CC FT MOD_RES 332 334 BY SIMILARITY.
CC FT MOD_RES 335 354 BY SIMILARITY.
CC FT MOD_RES 355 366 BY SIMILARITY.
CC FT MOD_RES 367 391 BY SIMILARITY.
CC FT MOD_RES 392 403 BY SIMILARITY.
CC FT MOD_RES 404 424 BY SIMILARITY.
CC FT MOD_RES 425 448 BY SIMILARITY.
CC FT MOD_RES 449 455 BY SIMILARITY.
CC FT MOD_RES 456 466 BY SIMILARITY.
CC FT MOD_RES 467 491 BY SIMILARITY.
CC FT MOD_RES 492 499 BY SIMILARITY.
CC FT MOD_RES 500 502 BY SIMILARITY.
CC FT MOD_RES 503 516 BY SIMILARITY.
CC FT MOD_RES 517 531 BY SIMILARITY.
CC FT MOD_RES 532 538 BY SIMILARITY.
```



```

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified, and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49212.6789948; -
CC HYPOTHETICAL PROTEIN; TRANSMEMBRANE; TRANSFERASE; GLYCOSYLTRANSFERASE.
FT FT TRANSMEM 337 357 POTENTIAL.
FT FT TRANSMEM 375 395 POTENTIAL.
FT FT TRANSMEM 415 435 POTENTIAL.
FT FT TRANSMEM 444 464 POTENTIAL.
FT FT TRANSMEM 508 528 POTENTIAL.
FT FT TRANSMEM 547 567 POTENTIAL.
FT FT TRANSMEM 572 592 POTENTIAL.
FT FT TRANSMEM 712 732 POTENTIAL.
FT FT TRANSMEM 1215 1235 POTENTIAL.
FT FT TRANSMEM 1268 1288 POTENTIAL.
FT FT TRANSMEM 1303 1323 POTENTIAL.
FT FT TRANSMEM 1370 1390 POTENTIAL.
FT FT TRANSMEM 1394 1414 POTENTIAL.
FT FT TRANSMEM 1475 1495 POTENTIAL.
FT FT TRANSMEM 1514 1534 POTENTIAL.
FT FT TRANSMEM 1549 1569 POTENTIAL.
FT FT TRANSMEM 1585 1605 POTENTIAL.
FT FT TRANSMEM 1615 1675 POTENTIAL.
FT FT TRANSMEM 1713 1733 POTENTIAL.
SQ SEQUENCE 1785 AA; 207482 MW; 4F171DF7 CRC32;

alignment_scores:
Quality: 61.00 Length: 48
Ratio: 2.259 Gaps: 2
Percent Similarity: 56.250 Percent Identity: 35.417

alignment_block:
US-09-030-606-115 x GLS3_YEAST ..

Align seg 1/1 to: GLS3_YEAST from: 1 to: 1785

187 AAAAAAATCTGCAAGAGCTAGTCTATCATGCAATGCAAGGTGAATTGCA 236
:::||||:|||| ||| :::||:||:||||| |||:::|||||
45 GlnlySerSerMetClnsMetIyrglnHisIeMetGlyGlnLeuS 61
237 T.....GTTCTCAGAACCATTTGCACGACAGCCT 268
| ||:::||||| ||:::|||||
61 pEarargAlaSerArgThrGlyAlaGlnHisAlaLeuValSerLeuHisV 78

269 GTTCTATCCTGTTTAATTAATTAAGTTGGGTTCTCTACATGCA 312
||| :::||||| |||||
78 alSerIyr.....IleGlyGlyGlnHisAla 86

seq_name: SwissProt_37:y056_YEAST

seq_documentation_block:
ID y056_YEAST STANDARD; PRT; 586 AA.
AC P46982;
DT 01-FEB-1996 (REL. 33, CREATED)
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 67.3 KD PROTEIN IN SW61-ATP12 INTERGENIC REGION.
DE YJ1186W OR J0409.
OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA: FUNGI: ASCOMYCOTA: HELMISCOMYCETES: SACCCHAROMYCETALES;
OC SACCCHAROMYCETACEAE: SACCCHAROMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RA OBERMAIER B., PIRAVANDI E., RINKE M., DOMDEY H.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- SIMILARITY: TO YEAST TTP1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: Z49461; G1008393; -  
KW HYPOThETICAL PROTEIN.  
SO SEQUENCE 586 AA; 67252 MW; B9D1BA39 CRC32:

alignment\_scores:  
Quality: 60.50 Length: 74  
Ratio: 1.287 Gaps: 3  
Percent Similarity: 63.514 Percent Identity: 27.027

alignment\_block:

US-09-030-606-115/rev x YUS6\_YEAST ..

Align seg 1/1 to: YUS6\_YEAST from: 1 to: 586

```

223 TCAGATGCTGATAGACTAGCTCTTCAGATGTTTCTACAC..... 182
||||| ||| : : : : : ||||| : :
479 SeraspLeuaspIlemePheIleIaIaSerPheYrLysAlaAspPr 495
181 .....TTCAGATGGGTATGATGATGTCATGCGGGAAGAAG 142
||||| : : : : : ||||| : :
495 OTrpPheUyrHisGluAsnArgPheIleGlyProAsnGlyGluGlnV 512
||||| : : : : : ||||| : :
141 CAGATGATGATCACAACCAAGTAATTTTAAACAACACATTTTGT 92
||||| : : : : : ||||| : :
512 aLArgGlyPheArgLysProHisArgYr.GlyMetAspPheGluLeuPh 528
||||| : : : : : ||||| : :
91 TTGCAACACATATACATACACAGTGAATGCTGATCTTCAATGCG 42
||||| : : : : : ||||| : :
528 eleuPheAsnAspMetArgGlySer...PheCysThrThrProLysSerG 544
||||| : : : : : ||||| : :
41 AAGTGAAGAATTAATTC 22
||||| : : : : : ||||| : :
544 InvalIleLysPheLysTyr 550

```

seq\_name: SwissProt\_37:VANH\_ENTFC

seq\_documentation\_block:

```

ID VANH_ENTFC STANDARD; PRT; 322 AA.
AC Q05709;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE D-SPECIFIC ALPHEA-KETO ACID DEHYDROGENASE (EC 1.1.1.-) (VANCOMYCIN
DE RESISTANCE PROTEIN VANH).
GN VANH.
OS ENTEROCOCCUS FAECIUM (STREPTOCOCCUS FAECIUM).
OG PLASMID PIP816.
OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; ENTEROCOCCACEAE;
OC ENTEROCOCCUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BM4147;
RX MEDLINE: 91348521.
RA ARTHUR M., MOLINAS C., DUTKA-MALEN S., COURVALIN P.;
RT "Structural relationship between the vancomycin resistance protein
RT VanH and 2-hydroxycarboxylic acid dehydrogenases.";
RL GENE 103:133-134(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BM4147; TRANSPOSON-TN1546;
RX MEDLINE: 93106944.
RA ARTHUR M., MOLINAS C., DEPARDEU F., COURVALIN P.;
RT "Characterization of Tn1546, a Tn3-related transposon conferring
RT glycopeptide resistance by synthesis of desipptide peptidoglycan
RT precursors in Enterococcus faecium BM4147.";
RL J. BACTERIOL. 175:117-127(1993).

```

```

RN [3]
RP SEQUENCE OF 1-32, AND CHARACTERIZATION.
RC STRAIN-BM4147;
RX MEDLINE: 92031480.
RA BUGG T.D.H., WRIGHT G.D., DUTKA-MALEN S., ARTHUR M., COURVALIN P.,
RA WALSH C.T.;
RT "Molecular basis for vancomycin resistance in Enterococcus faecium
RT BM4147: biosynthesis of a desipptide peptidoglycan precursor by
RT vancomycin resistance proteins VanH and Vana.";
RL BIOCHEMISTRY 30:10408-10415(1991).
RN [4]
RP FUNCTION.
RX MEDLINE: 92394903.
RA HANDWERGER S., PUCCI M.J., VOLK K.J., LIU J., LEE M.S.;
RT "The cytoplasmic peptidoglycan precursor of vancomycin-resistant
RT Enterococcus faecalis terminates in lactate.";
RL J. BACTERIOL. 174:5982-5984(1992).
CC -1- FUNCTION: REQUIRED FOR HIGH-LEVEL RESISTANCE TO GLYCOPEPTIDES
CC ANTIBIOTICS. CATALYSES THE REDUCTION OF 2-KETO ACIDS TO 2-D-
CC HYDROXY ACIDS THAT GIVE RISE TO PEPTIDOGLYCAN PRECURSORS THAT
CC TERMINATE IN THE DESIPEPTIDE D-ALANINE-2-LACTATE RATHER THAN
CC THE DIPEPTIDE D-ALANINE-D-ALANINE THUS PREVENTING VANCOMYCIN
CC BINDING.
CC -1- INDUCTION: BY VANCOMYCIN, MEDIATED BY VANS/VANR.
CC -1- SIMILARITY: BELONGS TO THE D-ISOMER SPECIFIC 2-HYDROXYACID
CC DEHYDROGENASES FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: M64304; G148330; -
DR EMBL: M97297; G155041; -
DR PIR: J01191; J01191.
DR PROSITE: PS00065; D_2_HYDROXYACID_DH_1; 1.
DR PROSITE: PS00670; D_2_HYDROXYACID_DH_2; 1.
DR PROSITE: PS00671; D_2_HYDROXYACID_DH_3; 1.
DR PFAM: PF00389; 2-Hacid_DH; 1.
DR HSSP: P26297; 1DLd.
DR OXIDOREDUCTASE: NAD; CELL WALL; ANTIBIOTIC RESISTANCE; PLASMID.
FT ACT_SITE 231 231 SUBSTRATE-BINDING (BY SIMILARITY).
SQ SEQUENCE 322 AA; 35795 MW; 0713B1F3 CRC32;

```

alignment\_scores:  
Quality: 60.00 Length: 88  
Ratio: 1.250 Gaps: 4  
Percent Similarity: 54.545 Percent Identity: 25.000

alignment\_block:

US-09-030-606-115/rev x VANH\_ENTFC ..

Align seg 1/1 to: VANH\_ENTFC from: 1 to: 322

```

276 ATAGAACAGGCTGTCTGCTGAATGTTCTGAGAACATCAAA...TT 230
::: ||||| : : : : : ||||| : :
152 ValGlyThrGlyGlnIleGlyLysAlaValIleGluArgLeuArgGlyPh 168
229 CACCTGTGATGATGCTGATGATGATGCTTTCAGATGTTTTCACCACTT 180
| ||| : : : : : ||||| : :
168 eGlyCysLysValLeu.....AlaTyrSerArgSerA 179
179 CAGAGATGGTAAATGACTAGTCCATGAGG.....AAAAGCA 139
::: : : : : : ||||| : :
179 rGserIleGluValAsnTyrValProPheAspGluLeuGluHisSer 195
138 GATGATTCACAACCAAGTAATTTTAAACAACACATTTTGT 89
||| ||| : : : : : ||||| : :
196 AspIleValThrLeuHisValProLeuAsnThrAspThrHisTyrIleI 212

```







This Page Blank (uspto)



405 yaspbys 407

seq\_name: sp.organelle:078246

seq\_documentation\_block:  
ID 078246 PRELIMINARY: PRT: 506 AA.

AC 078246;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)

DE MATURASE K.

GN MATK.

OS Attractylodes koreana.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Asteridae; euasterids II; Asterales; Asteraceae;

OC Attractylodes.

OC [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LEAF;

RX MEDLINE: 98297426.

RA MIZUKAMI H., SHIMIZU R., KOHJYOMA M., KOHDA H., KAWANISHI F.,

RA HIRAOKA N.;

RT "Phylogenetic analysis of Attractylodes plants based on chloroplast

trnK sequence."

RT Biol. Pharm. Bull. 21:474-478(1998).

DR EMBL: AB008760; BAA32657.1; -

DR PFAM: PF01348; Intron\_maturas2; 1.

KW Chloroplast.

SQ SEQUENCE 506 AA: 59636 MW; 18F5472D CRC32;

alignment\_scores:

Quality: 68.50 Length: 74

Ratio: 1.631 Gaps: 5

Percent Similarity: 56.757 Percent Identity: 33.784

alignment\_block:

US-09-030-606-115/rev x 078246 ..

Align seg 1/1 to: 078246 from: 1 to: 506

238 CATCAATTCACCTG.....TC 222

152 TyrProAlaHisLeuGluLeuValGlnAlaLeuArgTyrTrpIleG1 168

221 AGATGCTGATAGACTAGCTCTTCAGATGTTTTC.....TACCACT 181

168 uAspAlaSerSerLeuHisLeuArgPhePheLeuHisGluTyrHisA 185

180 TCAGAGATGGGTTAATGACTAGTTCGAATGGGAAAGCAAGATGAGAT 131

185 snArgAspSerLeuIleThr...SerAsnSerLysLysAla..... 197

130 CACAACCAAGTAATTTTAACAAGACAC.....TTTTTTTTTT 90

198 .....SerSerSerPheSerLysArgAsnHisArgLeuPhePhe 212

89 TGCACACACATATATACATCACAG 68

212 uTyrThrSerHisValCysGlu 219

seq\_name: sp.organelle:079983

seq\_documentation\_block:

ID 079983 PRELIMINARY: PRT: 506 AA.

AC 079983;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)

DE MATURASE K.

GN MATK.

OS Attractylodes ovata, Attractylodes chinensis, Attractylodes japonica,

OS and Attractylodes lancea.

OC Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;

OC core eudicots; Asteridae; euasterids II; Asterales; Asteraceae;

OC Attractylodes.

OC [1]

RP SEQUENCE FROM N.A.

RC TISSUE-LEAF;

RX MEDLINE: 98297426.

RA MIZUKAMI H., SHIMIZU R., KOHJYOMA M., KOHDA H., KAWANISHI F.,

RA HIRAOKA N.;

RT "Phylogenetic analysis of Attractylodes plants based on chloroplast

trnK sequence."

RT Biol. Pharm. Bull. 21:474-478(1998).

DR EMBL: AB008760; BAA32657.1; -

DR PFAM: PF01348; Intron\_maturas2; 1.

KW Chloroplast.

SQ SEQUENCE 506 AA: 59663 MW; 586D308D CRC32;

alignment\_scores:

Quality: 68.50 Length: 74

Ratio: 1.631 Gaps: 5

Percent Similarity: 56.757 Percent Identity: 33.784

alignment\_block:

US-09-030-606-115/rev x 079983 ..

Align seg 1/1 to: 079983 from: 1 to: 506

238 CATCAATTCACCTG.....TC 222

152 TyrProAlaHisLeuGluLeuValGlnAlaLeuArgTyrTrpIleG1 168

221 AGATGCTGATAGACTAGCTCTTCAGATGTTTTC.....TACCACT 181

168 uAspAlaSerSerLeuHisLeuArgPhePheLeuHisGluTyrHisA 185

180 TCAGAGATGGGTTAATGACTAGTTCGAATGGGAAAGCAAGATGAGAT 131

185 snArgAspSerLeuIleThr...SerAsnSerLysLysAla..... 197

130 CACAACCAAGTAATTTTAACAAGACAC.....TTTTTTTTTT 90

198 .....SerSerSerPheSerLysArgAsnHisArgLeuPhePhe 212

89 TGCACACATATATACATCACAG 68

212 uTyrThrSerHisValCysGlu 219

seq\_name: sp.invertebrate:Q17945

seq\_documentation\_block:

ID Q17945 PRELIMINARY: PRT: 462 AA.

AC Q17945; Q22858;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-JAN-1999 (TREMBLrel. 09, last annotation update)

DE T28H10.3 PROTEIN.

GN T28H10.3.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

OC Rhabditina; Rhabditidae; Rhabditidae; Pelodierinae; Caenorhabditis.

OC [1]

RP SEQUENCE FROM N.A.

RA DOBSON R.;

KL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

NP [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
RA BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,  
RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,  
RA GADNER A., GREEN P., HAWKINS T., HILLER L., JIER M., JOHNSON L.,  
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LARRELLE P.,  
RA LIGHTING J., LOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
RA THERIY-NEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
RA MATSON A., WEINSTOCK L., WILKINSON-SPRAT J., WOLDMAN P.,  
RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.\*  
RL Nature 368:32-38(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA KERSHAW J.,  
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL: 277653; CAB01126.1; -  
DR EMBL: 275551; CAB01126.1; JOINED.  
DR EMBL: 275551; CA99935.1; -  
DR EMBL: 277653; CA99935.1; JOINED  
SO SEQUENCE 462 AA: 53237 MW: D9A96DC4 CRC32:

alignment_scores:	
Quality:	66.50
Ratio:	84
Percent Similarity:	59.524
Percent Identity:	23.810

alignment\_block:  
US-09-030-606-115/rev x Q17945 .

Align seg 1/1 to: Q17945 from: 1 to: 462

298 CCAAACTATTATTATTAACAGGATGAAACAGCGCTGTGGGTCAAAATG 24.9  
124 ProGluAsnPheLeuAsnValLeuIysGlyAsnAlaSerGlyIleAspG1 14.0  
248 TTCTGAGAACCATCCACATTCACCTGCAGATGCTGATAGACTAGCTTTC 19.9  
140 yGcIyaSncIyArGValLeuGluTUrAsnAspAsnAspArg..... 15.3  
198 AGATGTTTTCTAC.....CAGTTC 17.9  
154 ..ValPheValIyTrPheTrAspPhISgIyAlaValGlyMetIleSerPhe 16.6  
178 AGAGATGGGTTATGACTAGTTCAAAT...GGGAAAAACCAAGATGAT 13.2  
170 ProAspGlyIleLeuThrValIysGlnLeuAsnAspValLeuValTrpMe 18.6  
131 TCACAAACCAAGTAATTTAAACAAAGACACTTTTTTTTTTGGCAACAC 8.2  
186 tHISyAsnIbS/yLSyTyrSerGlnLeuThrPheTyrIleuGluIbAcSg 20.3

seq\_name: sp\_Invertebrate:045162

seq_documentation_block:	seq_documentation_block:
ID 045162	PRELIMINARY; PRT; 493 AA.
AC 045162:	
DT 01-JUN-1998 (TREMBLrel_06, Created)	
DT 01-JUN-1998 (TREMBLrel_06, Last sequence update)	
DT 01-MAY-1999 (TREMBLrel_10, Last annotation update)	
DE F37B4.5 PROTEIN.	
DE F37B4.5.	
GN	
OS <i>Caenorhabditis elegans</i> .	
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida	

OC Rhabdittina; Rhabdittidae; Rhabdittidae; Peloderiinae; Caenorhabdittis.  
 RP [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BORTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU Z., DUBBIN R., FAVELLO A., FULTON L.,  
 RA CARDNER A., GREEN P., HARRIS T., HILLER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATRILLE P.,  
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIKEN L., ROOPA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STULSTON J.,  
 RA THIERY-IEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLMAN P.;  
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RL *elegans*.  
 RL Nature 368:32-38(1994).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA HENKHAUS J., WOHLDMANN P., O'BRIEN D.;  
 RN Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RP [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF047657; AAC04409.1; -.  
 QO PPRM; PF01461; 7tm\_4; 1.  
 QO SEQUENCE 493 AA; 5719 MW; 4C27B2FD CRC32;

```

alignment_scores:
  Quality: 66.00      Length: 133
  Ratio: 0.985      Gaps: 7
Percent Similarity: 50.376      Percent Identity: 26.316

```

```
alignment_block:
US-09-030-606-115 x 045162 .
```

Align seg 1/1 to: 045162 from: 1 to: 493

```

5  TTCTCTCCSCCTCCCTGAAATTAATTAATTTTCCAACTGGCAATTTGCAAG 54
16 PheLeuAsnHisIaIaPheAsn.....LeuThrTrpCysLeuGlnIlePr 30
55 ATATACACATT...TCACCTGATGATATATGT.....T 86
30 oHeHisIleLeuGlyAlaTyCysIleLeuCysIlyThrProAspSerM 47
87 GCACAAAAAAGTCTCTTGGTT..... 112
47 etySseValIlyStrpSerMetPheAsnLeuHisPheTrpSerValLeu 63
113.....AAATATCTTGTTGTGTAATCCATCTGCTTTTCCCAT 153
64 LeuAspLeuIlySleThrPheLeuIleSerProPheValLeuPheProAl 80
154 TGGAACTAGCATTAACCCATCTCGAACTGGTAGAAAAACATCTGAGA 203
80 aPheAlaGlyPhe...ProLeuGlyVal.....LeuIlyst 91
204 GCTAGCTATTCAGACATCTACAGCGGAATGGATGGTTCTCAGAACATT 253
91 yThrTrpIlyIleSerIleAspValGlnThrTyrlLeuIleLeuIleSerTy 107
254 TCACCC...AGACAGCTGTGTTCTATCTCGTTTATAAATATGTTGGGT 300
108 SerAlaLeuGlySerSerIleLeuThrLeuPheGlnAsnAlaGlyTyrlPhe 124
301 TCCTCATGATGATACAAACCTGCTGCCAATCTGTGCACATAAAGTGTG 349

```



RT elegans.":  
 RL Nature 368:32-38(1994).  
 DR EMBL: 219157; CAA79568.1;  
 SO SEQUENCE 826 AA; 91638 MW; D7B8FF0E CRC32;

alignment\_scores:  
 Quality: 65.00 Length: 103  
 Ratio: 1.226 Gaps: 6  
 Percent Similarity: 51.456 Percent Identity: 33.981

alignment\_block:  
 US-09-030-606-115/rev x P91845 ..

Align seg 1/1 to: P91845 from: 1 to: 826

```

291 AATTATTAACAGATAGAACAGCTGTCTGGTGAATGCTTCTGAG 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
322 AsnGlyLeuAsnArgLeuLeuSerAspAlaSerAsnThrMetValLeuG1 338
242 .....AACCATCCAAATTCACCTGTGAGATGCTG 214
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
338 uUeYrSerIlePheValArgThrHisSerArgLeuSerGluTrpL 355
213 ATAGACTAGCTCTTCAGATGTTTCTTACCACTGACAGATGGTTAATG 164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
355 euArgLeuAlaLeuAlaLysLeuPhe.....AlaArgLysAlaLagL 369
153 ACTAGTTCGAATGGGAAAAGCAAGATGATTCACAAACCAAGTAATT 114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
370 ThrLeuProAsnThrLysLysGlnIleGlyHisThrLeuAsnValIleL 386
113 TAAACAAGACACTTTTCTTTTTCGAACACATATACATCAC..... 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
386 eu.....GluCysPhe.....AsnAlaHisHisGlnLeu 395
69 .....AGTGAATGTGTATCTTCGCAATTCGCAAGTTGAAG 32
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
396 ValThrValCysGluLeuMetCysAspProIleHisLeuMetValProL 412
31 AATTAAA 25
    : : : : :
412 SALAArG 414

```

seq\_name: sp\_invertebrate:001744

seq\_documentation\_block:  
 ID 001744 PRELIMINARY; PRT; 347 AA.

AC 001744;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)  
 DE COSMID F35F10.  
 GN F35F10.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;  
 OC Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RX MEDLINE: 94150718.  
 RA WILSON R., AINSOUCO R., ANDERSON K., BAYNES C., BERKS M.,  
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,  
 RA CRAXTON M., DEAR S., DU 2., DURBIN R., FAVELLO A., FULTON L.,  
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,  
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,  
 RA LIGHTING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,  
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,  
 RA SMAILDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,  
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,  
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";

RL Nature 368:32-38(1994).  
 RN [2]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-BRISTOL N2;  
 RA ROHLFING T.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF002198; AAB53892.1;  
 SO SEQUENCE 347 AA; 39433 MW; 2DF40D93 CRC32;

alignment\_scores:  
 Quality: 64.50 Length: 135  
 Ratio: 1.040 Gaps: 6  
 Percent Similarity: 45.926 Percent Identity: 24.444

alignment\_block:  
 US-09-030-606-115 x 001744 ..

Align seg 1/1 to: 001744 from: 1 to: 347

```

1 GCTCTTCTCCCTCCCTGTAATTAATCTTCACTGCAATTTGC 50
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
167 SerLeuThrIleGluLysProGlnPheAsnLysPhePheMetGlnIleG1 183
51 AAGATTACACATTTCACTGTGATGATATATGTTGCAAAAAA 100
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
183 uLeuLysAsnIleLysValPheLysThrGluLeuAsnAsnArgSers 200
101 GTCGTCTTGTTAATAATTAATCTTGTGTAATCCATCTTCTTTCC 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 erVal.....GluSerIleSerIleValVal 208
151 CATTCGACATGCTCA..... 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
209 TyrTrpAsnIleLysLysAlaSerSerLeuLeuGluPheSerLysSerCy 225
166 .....TTAACCATCTCTGAACGTGTAGAAAAACAT 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
225 sAsnAlaMetGlnAsnCysLeuAspAlaLeuArgCysGlnGlyThrG 242
197 CTGAAGAGCTACTCTATCAGCATCTGACAGGTGAATTTGGATGTTCTAG 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
242 LuHisLeuAlaMetLysGluLysMetArgGlyGlnCys.....ValThr 256
247 AACCATTTACCCAGACAGCGTGTCTATCTGCTGTTAATAATTAAGTTT 296
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
257 MetHisPhePheSerThrAspValPhe...SerCysLeu..... 268
297 GCGTCTCTCATGCATCAACAAC.....CCTGCTCCCAATCTGCATC 340
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
269 .....ThrLysLeuGlnAsnSerLysProLysProGluCysTyG1 282
341 AAAA 344
    |||
282 uLys 283

```

seq\_name: sp-organelle:033098

seq\_documentation\_block:  
 ID 033098 PRELIMINARY; PRT; 353 AA.

AC 033098;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
 DE NATUREASE (FRAGMENT).  
 GN MATK.  
 OS Saxifraga ferruginea.  
 OG Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Saxifragales; Saxifragaceae; Saxifraga.

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-LEAF;  
 RA JOHNSON L.A., SOLTIS D.E.;  
 RL Ann. Mo. Bot. Gard. 0:0-0(0).  
 DR EMBL: L34141; AAA84602.1;  
 DR MENDEL, 5094; SAKFE:ycf14.1.  
 KW Chloroplast.  
 FT NON\_TER 353  
 SQ SEQUENCE 353 AA: 42259 MW: 46AA38A9 CRC32;

alignment\_scores:  
 Quality: 64.50 Length: 71  
 Ratio: 1.743 Gaps: 3  
 Percent Similarity: 52.113 Percent Identity: 30.986

alignment\_block:  
 US-09-030-606-115/rev x Q33098 ..  
 Align seg 1/1 to: Q33098 from: 1 to: 353

238 CATCCATTACCTG.....TC 222  
 154 HSPGTEHSLLEuValGlnThrLeuArgTyrArgVally 170  
 221 AGATGCTGATGACTGCTTCAGATGTTTC.....TACAGT 181  
 170 SASPAlaserSerLeuHSLLeuArgPheLeuTyrGluTyrGlnA 187  
 180 TCAGAGAGCGTTATGCTACTGTCATCGGGAAGCAAGATGAT 131  
 187 snArgasnSerLeuThrProLysLysGluLeuSerLeval 203  
 130 CACAACCACTTAATTTAAACAAGACACTTTTTTTTTCACACA 81  
 204 SerLysArgasn.....GlnArgLeuHeuPheLeuTyrAsnT 217  
 80 ATATACATCACAG 68  
 217 RTyrValCysGlu 221

seq\_name: sp.archea:059203

seq\_documentation\_block:  
 ID 059203 PRELIMINARY; PRT; 136 AA.  
 AC 059203;

DT 01-AUG-1998 (TREMBlrel. 07, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
 DE 136AA LONG HYPOTHETICAL PROTEIN.

GN PH1534.

OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN-OT3;

RX MEDLINE: 9834437.

RA KAMARABAYASI Y., SAMADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,  
 RA YAMAMOTO S., SEKINE M., BABA S., KOSUGI H., HOSOVAMA A., NAGAI Y.,  
 RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMIYA M., OHFUKU Y.,  
 RA FUNAHASHI T., TANAKA T., KUDOH T., YAMAZAKI J., KUSHIDA N., OGUCHI A.,  
 RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,  
 RA KIKUCHI H.

RT "Complete sequence and gene organization of the genome of a  
 hyper-thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
 DR DNA Res. 5:55-76(1998).

EMBL: AP00006; BAA30644.1; -  
 DR EMBL: AP00006; BAA30644.1; -  
 SQ SEQUENCE 136 AA: 15267 MW: 5F920399 CRC32;

alignment\_scores:  
 Quality: 63.50 Length: 64  
 Ratio: 1.671 Gaps: 2  
 Percent Similarity: 59.375 Percent Identity: 31.250

alignment\_block:  
 US-09-030-606-115 x 059203 ..

Align seg 1/1 to: 059203 from: 1 to: 136

147 TCCCATTTGAGTACTGATTAACCATCT.....CTGACTGTGA 187  
 25 ThrProValGluThrAlaIleHisProSerPheSerLeuLeuIstRpa1 41  
 188 GAAAACATCTGAAGCTAGTCTA.....TAGCATCTGACAGGTGA 231  
 41 aserAsnIleaspProserAsnIleArgProserSerIleaspLysIle1 58  
 232 TTGGATGTTCTCAGAACCATTCACACAGACGCTGTTTATCCTGT 281  
 58 legLypPheThrProValGlnThrSerProLeuGlyThrThrMetHisAla 74  
 282 TTATAAATTAGTTGGGCTCTCTACATGCATTAACAACCCCT 323  
 75 SerLeuSerMetGlyGlyPheThrIleLeuIleThrAsnPro 88

seq\_name: sp.organelle:Q31804

seq\_documentation\_block:  
 ID Q31804 PRELIMINARY; PRT; 352 AA.  
 AC Q31804;

DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)  
 DE MATURASE (FRAGMENT).

GN MATK.

OS Astilbe japonica x Astilbe chinensis.

OC Chloroplast.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
 OC core eudicots; Saxifragales; Saxifragaceae; Astilbe.

RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE-LEAF;

RA JOHNSON L.A., SOLTIS D.E.;

RL Ann. Mo. Bot. Gard. 0:0-0(0).

DR EMBL: L34114; AAB03515.1;  
 DR MENDEL, 5033; ASTJ:ycf14.1.

KW Chloroplast.

FT NON\_TER 352  
 SQ SEQUENCE 352 AA: 41793 MW: B399E1B8 CRC32;

alignment\_scores:  
 Quality: 63.50 Length: 71  
 Ratio: 1.512 Gaps: 4  
 Percent Similarity: 59.135 Percent Identity: 29.577

alignment\_block:  
 US-09-030-606-115/rev x Q31804 ..

Align seg 1/1 to: Q31804 from: 1 to: 352

238 CATCCATTACCTG.....TC 222  
 153 HSPGTEHSLLEuValGlnThrLeuArgTyrTrpVally 169  
 221 AGATGCTGATGACTA...GCTCTCAGATGTTTCTTACCACTTGA 176  
 169 SASPAlaserSerLeuHSLLeuArgPhePheTyrGluTyrGlnA 186  
 175 .....GATGGTTAATGACTAGTTCATATGGGGAAGCAAGATGAT 131



```

186 sncysasSerleulerProtnrPolyslySerleleval 202
130 CACAACCAAGTAATTTAAACAAAGACCTTTTTCGACACA 81
203 SerlysaIgaasn.....GlnrlygLeupheuleuTyraIga 216
80 ATATACATCAGC 68
216 rTyraIgaSerlu 220

```

seq\_name: sp\_plant:024093

```

seq_documentation_block:
ID 024093 PRELIMINARY: PRT: 569 AA.
AC 024093:
DT 01-JAN-1998 (TEMBLrel. 05, Created)
DT 01-JAN-1998 (TEMBLrel. 05, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE L-ASCORBATE OXIDASE PRECURSOR.
GN MTN23.
OS Medicago truncatula (barrel medic).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
OC Medicago.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CV. JEMALONG J5; TISSUE-ROOT MODULE;
RX MEDLINE: 96212994.
RA GAMAS P.; DE CARVALHO NIEBEL F.; LESCURE N.; CULLIMORE J.;
RT "Use of a subtractive hybridization approach to identify new Medicago
RL M01. Plant Microbe Interact. 9:233-242(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. JEMALONG J5; TISSUE-ROOT MODULE;
RX GAMAS P.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: Y15295; CA75577.1;
DR FRAM: PF00394; Cu-oxidase; 2.
DR PROSITE: PS00079; MULTICOPPER_OXIDASE1; 1.
DR PROSITE: PS00080; MULTICOPPER_OXIDASE2; 1.
KM Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 569 POTENTIAL.
SQ SEQUENCE 569 AA: 63973 MW: 7679894D CRC32:

```

alignment\_scores:

Quality:	63.50	Length:	40
Ratio:	2.646	Gaps:	1
Percent Similarity:	60.000	Percent Identity:	45.000

alignment\_block:

us-09-030-606-115/rev x 024093

```

Align seg 1/1 to: 024093 from: 1 to: 569
286 ATTAACAGATAGAACAGGCTGTCTGGTGAATGTTCTGACAAACA 237
453 ILeuGlnAaenAlaAnGlnLeuAnGly..AsnGlySerGluIleH1 468
236 TCCAAATTCACCTGCATGCTGATAGACTAGCTCTTCAGATGTTTCT 187
468 sProtrpHisLeuHisGlyHisAspSerpTrpValLeuGlyTyGlyGluG 485
186 ACCAGTTCAGAGATGGTTA 167
485 lYArGpHeArGpProGlyVal 491

```

seq\_name: sp\_rudent:0922E3

```

seq_documentation_block:
ID 0922E3 PRELIMINARY: PRT: 911 AA.
AC 0922E3:
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TEMBLrel. 10, Last annotation update)
DE IREL.
GN IREL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN (1)
RP SEQUENCE FROM N.A.
RA WANG X.Z.; HARDING H.P.; ZHANG Y.; JOLICOEUR E.M.; KURODA M.; ROW D.;
RT "Cloning of mammalian Irel reveals a diversity in the ER stress
RL responses."
RL EMBO J. 17:5708-5717(1998).
DR EMBL: AF071777; AAC64400.1;
SQ SEQUENCE 911 AA: 101355 MW: 57240D23 CRC32:

```

alignment\_scores:

Quality:	63.50	Length:	114
Ratio:	0.992	Gaps:	5
Percent Similarity:	56.140	Percent Identity:	25.439

alignment\_block:

us-09-030-606-115 x 0922E3

```

Align seg 1/1 to: 0922E3 from: 1 to: 911
4 CTTCTCCCTCCCTCGAATTAATCTTCAACTGCAATTCGAC 53
262 lIeArGLeuProAlaSerSerTyGlnsPthrAlaThrGlnPheSer 278
54 GATTACACATTTCACTGATGATATATGTTGTCACAAAAAAGTG 103
278 rLeuAspThrGlnLeuLeuMetThrLeuTyValGlyLysGluGluAlaG 295
104 TCTTT...GTTAAATTAATCTGTTGTTGCAATTCCTTTTCCC 150
295 lYpHeTyValSerLysAlaLeuValHisAlaGlyValAlaLeuValPro 311
151 CATGGAACTAGTCATTAACCCATCTCGAACSTGTAGAAAAACATCGA 200
312 ArG.....GlyLeuThrLeuAlaProMetAspGlyProThrThrAs 325
201 AGAGCTAGTCTATCAGCATCTGACAGGATGATGATGTTCTCGAAC 250
325 pGluValThrLeuGln..ValSerGlyGluArgGluGlySerPro.Ser 340
251 ATTTCACCCAGACGCTGTTCTATCCTGTTTAATAATTAAGTTGGGT 300
341 ThrAlaValArgTyProSerGlySerValAlaLeuProSerGlnTrpLe 357
301 TCTCTACATGATACAAACCCCTGCTCAATCTGTAC 338
357 uLeuIleGlyTyThrHisGlu.ProProProValLeuHis 369

```

seq\_name: sp\_bacteria:046722

```

seq_documentation_block:
ID 046722 PRELIMINARY: PRT: 421 AA.
AC 046722:
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-NOV-1996 (TEMBLrel. 08, Last annotation update)
DE HYPOHETICAL 47.2 KD PROTEIN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN (1)
RP SEQUENCE FROM N.A.

```

RC STRAIN-011, M92;  
RX MEDLINE: 96060831.  
RA BASTIN D.A.; REEVES P.R.;  
RT "Sequence and analysis of the O antigen gene (rfb) cluster of  
RT Escherichia coli O11.";  
RL Gene 164:17-23(1995).  
DR EMBL: U13629; AAC44886.1; .  
KW Hypothetical protein.  
SO SEQUENCE 421 AA; 47202 MW; 2699342D CRC32;

alignment\_scores:  
Quality: 63.00 Length: 85  
Ratio: 1.465 Gaps: 5  
Percent Similarity: 50.588 Percent Identity: 28.235

alignment\_block:  
US-09-030-606-115 x 046722 ..

Align seg 1/1 to: 046722 from: 1 to: 421

```
116 ATTAAGTGGTTGGATGATCCATCTGCTTTCCCATGGAAGACTAGTCA 165
    ||| ||||| ||||| ||||| |||||
21 IleGlnPheValAsnProIleCysIlePhe..... 31
166 TTAACCATCTCTGAACTGTAGAAAACATCTGAGAGCTAGTCTATCA 215
    ||||| ||||| ||||| ||||| |||||
32 ....IleIleThrProLeuIleLeuAsnHisLeuGlyLysGlnSerTyrG 47
216 GCATCTGACAGGTGAATGATGTTCTCAGAACCATTTCAACCCAGACAG 265
    :: ||||| ||||| ||||| |||||
47 Ly.....AsnTrpIleLeuIleThrIle..... 55
266 CCTGTTCTATCCTTTAATAATTAAGT.....TG 297
    ::|||::|||::|||::|||::|||
56 .....ValSerPheSerGlnLeuIleCysGlyGlyCysSerAlaTr 69
298 GGT.....CTCTACATGATTAACAACCTGCTCAATCTGTCAATCA 341
    ||::|||::|||::|||::|||::|||
69 PileAlaLysIleIleIleAlaGlnGlnArgIleLeuSerAspLeuSerLysL 86
342 AAGT 346
    ||::|||
86 ysAsn 87
```

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:25 ; Search time 1809.22 Seconds  
(without alignments)  
399.038 Million cell updates/sec

Title: US-09-030-606-115  
Perfect score: 366  
Sequence: 1 GCCTTCTCTCCCTCCCTC.....CTGACTGAGTTAGTC.366

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	365	99.7	455	40	AA972883	OP24G07.s
2	365	99.7	538	49	AI620178	AI620178.tu54a11.x
3	363.4	99.3	435	38	AA809587	AA809587.n143e11.s
4	354	96.7	599	39	AA813266	AA813266.a144g03.s
5	352	96.2	490	33	AA437224	AA437224.zv54b10.s
6	342	93.4	423	50	AI685682	AI685682.t89h03.x
7	313	85.5	419	30	AA229495	AA229495.nc35g03.s
8	304.4	83.2	411	36	AA640928	AA640928.m28d08.r
9	299	81.7	383	50	AI672753	AI672753.w58d10.x
10	294.2	80.4	445	34	AA533772	AA533772.n193c05.s
11	292	79.8	423	34	AA531606	AA531606.n165e04.s
12	288	78.7	403	34	AA493522	AA493522.n975f05.s
13	284.2	77.7	339	36	AA650104	AA650104.n892f11.s
14	281	76.8	377	50	AI669511	AI669511.wb8f07.x
15	280	76.5	374	26	AA7380	AA7380.zc39h11.s1
16	280	76.5	356	34	AA507804	AA507804.n989g12.s
17	279	76.2	366	46	AI424280	AI424280.t81d02.x
18	278	76.0	386	41	AI027196	AI027196.ov83g09.x
19	273	74.6	380	39	AA833543	AA833543.a136b07.s
20	271.8	74.3	382	50	AI692186	AI692186.wd37g02.x
21	270	73.8	371	38	AA758204	AA758204.ab68e11.s
22	266.4	72.8	353	34	AA508806	AA508806.nb69c06.s
23	254.4	69.5	354	36	AA603372	AA603372.np06d01.s
24	248.8	68.0	337	33	AA400418	AA400418.zu69g10.s
25	241.8	66.1	282	36	AA654919	AA654919.n177e01.s
26	241.4	66.0	342	30	AA228941	AA228941.nc35h08.s
27	225	61.5	472	36	AA640241	AA640241.m20h08.s
28	223.6	61.1	261	36	AA650230	AA650230.n893g02.s
29	199.8	54.6	311	34	AA527743	AA527743.n958e05.s
30	192.4	52.6	346	50	AI701376	AI701376.wd91f07.x
31	172	47.0	283	36	AA652376	AA652376.n52h09.s
32	159	43.4	267	35	AA572913	AA572913.mn42f12.s
33	102.2	27.9	339	40	AA917791	AA917791.om82c01.s
34	100	27.3	196	35	AA579219	AA579219.nf36g08.s
35	99	27.0	215	43	AI201254	AI201254.qf70h05.x
36	92	25.1	245	36	AA622798	AA622798.np77b06.s
37	90	24.6	157	43	AI218326	AI218326.qh16d03.x
38	90	24.6	324	46	AI423843	AI423843.tg35c04.x
39	89	24.3	292	50	AI695410	AI695410.w679g06.x
40	88	24.0	415	41	AI052805	AI052805.o211f08.x
41	87	23.8	543	34	AA528104	AA528104.n158e05.s
42	87	23.8	504	34	AA531255	AA531255.n109a12.s
43	87	23.8	376	37	AA677355	AA677355.z158e10.s
44	87	23.8	414	42	AI139031	AI139031.qc22e12.x
45	87	23.8	277	46	AI424769	AI424769.t099g07.x

#### ALIGNMENTS

RESULT 1  
LOCUS AA972883 455 bp mRNA  
DEFINITION OP24G07.S1 Soares\_NFL\_T-GBC\_S1 Homo sapiens cDNA clone  
IMAGE:1577820 3', mRNA sequence.  
ACCESSION AA972883  
NID 93148063  
VERSION AA972883.1 GI:3148063

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 453)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 9, 1998 this sequence version replaced gi:937931.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 633 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 443.  
Location/Qualifiers  
1..455  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1577820"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled. Vector: pT7T3D-Pac (Pharmacia) with  
a modified polylinker. Site\_1: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19W, testis NHT, and B-cell  
NCI-CGAP GCBI) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."

BASE COUNT 162 a 68 c 94 g 131 t

ORIGIN

Query Match 99.7%; Score 365; DB 40; Length 455;

Best Local Similarity 100.0%; Pred. No. 8.7e-74;

Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTCCTTCTCCCTCCTCTGAAATTTAATCTTCAACTGCAATTGCAAGATTACAC 61  
|||||  
DB 453 CTCCTTCTCCCTCCTCTGAAATTTAATCTTCAACTGCAATTGCAAGATTACAC 394  
QY 62 ATTACAGTATGATATATGTTGGCAAAAAAAGTCTTTGTTTAAATTACT 121  
|||||  
DB 393 ATTCACTGATATATGTTGGCAAAAAAAGTCTTTGTTTAAATTACT 334  
122 TGGTTTGAATCCATCTGCTTTTCCCATTTGGAATGATCTATTAACCATCTCGAA 181  
|||||  
DB 333 TGGTTTGAATCCATCTGCTTTTCCCATTTGGAATGATCTATTAACCATCTCGAA 274  
QY 182 CTGTAGAAAAACATCTGAAGAGTAGTCTATAGCATCTGACAGGTGAATTGGATGTT 241  
|||||  
DB 273 CTGTAGAAAAACATCTGAAGAGTAGTCTATAGCATCTGACAGGTGAATTGGATGTT 214  
QY 242 CTCGAACCATTTACCCAGACGCTGTTCTATCCCTTTAATTAATTGTTGGTT 301  
|||||  
DB 213 CTCGAACCATTTACCCAGACGCTGTTCTATCCCTTTAATTAATTGTTGGTT 154  
QY 302 CTCACATGATTAACAAACCTGCTCCATCTGCATATAAAGTGTGACTGAAGTT 361  
|||||  
DB 153 CTCACATGATTAACAAACCTGCTCCATCTGCATATAAAGTGTGACTGAAGTT 94  
QY 362 TAGTC 366  
|||||

DB 93 TAGTC 89

RESULT 2  
A1620178/c  
LOCUS  
DEFINITION  
tuf5a1l.x1 NCI-CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2254844 3',  
mRNA sequence.  
ACCESSION  
A1620178  
NID  
g4628304  
VERSION  
A1620178.1 GI:4629304  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 538)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Mar 10, 1998 this sequence version replaced gi:2949075.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/db/ftp/image/image.html](http://www.bio.lnl.gov/db/ftp/image/image.html)

FEATURES  
source

Seq primer: -40UP from G1bco  
High quality sequence stop: 456.  
Location/Qualifiers

1..538  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2254844"  
/clone\_lib="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate. Vector: pT7T3D-Pac (Pharmacia)  
with a modified polylinker. Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonoids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 190 a 84 c 113 g 151 t

ORIGIN

Query Match 99.7%; Score 365; DB 49; Length 538;

Best Local Similarity 100.0%; Pred. No. 8.5e-74;

Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTCCTTCTCCCTCCTCTGAAATTTAATCTTCAACTGCAATTGCAAGATTACAC 61  
|||||  
DB 450 CTCCTTCTCCCTCCTCTGAAATTTAATCTTCAACTGCAATTGCAAGATTACAC 391  
QY 62 ATTCACTGATATATGTTGGCAAAAAAAGTGTGTTTAAATTACT 121  
|||||  
DB 390 ATTCACTGATATATGTTGGCAAAAAAAGTGTGTTTAAATTACT 331  
QY 122 TGGTTTGAATCCATCTGCTTTTCCCATTTGGAATGATCTATTAACCATCTCGAA 181  
|||||

Db 330 TGGTTGTGATCATTCTCTTTTCCCATTTGGAACATGATTAACCATCTCGAA 271  
OY 182 CTCGTAGAAAAACATCGAAGAGCTAGTCTATGACATCTGACAGTGAATGGATGTT 241  
Db 270 CTGTGAAAAAATCTGTAAGAGCTAGTCTATGACATCTGACAGTGAATGGATGTT 211  
OY 242 CTCGAAACATTTCCACCAGACACCTGTTTCTATCCTGTTTAATTAATTAGTTGGTT 301  
Db 210 CTCGAAACATTTCCACCAGACACCTGTTTCTATCCTGTTTAATTAATTAGTTGGTT 151  
OY 302 CTCATCATGATTAACAACCCCTGCTCCATCTGTACATAAAGCTGTGACTTGAAGTT 361  
Db 150 CTCATCATGATTAACAACCCCTGCTCCATCTGTACATAAAGCTGTGACTTGAAGTT 91  
OY 362 TAGTC 366  
Db 90 TAGTC 86

RESULT 3  
LOCUS AA809587 435 bp mRNA EST 18-FEB-1998  
DEFINITION n143611.s1 NCI\_CGAP\_Pr9 Homo sapiens CDNA clone IMAGE:595276, mRNA  
ACCESSION AA809587  
NID 92878993  
VERSION AA809587.1 GI:2878993  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 435)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Jan 19, 1998 this sequence version replaced gi:2287384.

JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 605 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from AmerSham.  
Location/Qualifiers  
1. .435  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="902E09; 2: 2p11.2-2p13.1"  
/clone="IMAGE:595276"  
/clone\_1lb="NCI\_CGAP\_Pr9"  
/sex="male"  
/tissue\_type="normal prostatic epithelial cells"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: PAMP10; mRNA made from  
normal prostatic epithelial cells, cDNA made by oligo-dT  
priming. Non-directionally cloned. Size-selected on  
agarose gel, average insert size 600 bp. Library made by  
D. Kitzman, NIH"

BASE COUNT 120 a 96 c 68 g 151 t  
ORIGIN

Query Match 99.3%; Score 363.4; DB 38; Length 435;

Best Local Similarity 99.7%; Pred. No. 2e-73;  
Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CTCCTTCTCCCTCCTCTGGAATTTATCTTTCACTTGCAATTTGCAAGATTACAC 61  
Db 21 CTCCTTCCCTCCCTCCTCTGGAATTTATCTTTCACTTGCAATTTGCAAGATTACAC 80  
OY 62 ATTTCACTGTGATGATATGTTGTCGAAAAAATAAGTCTTTGTTAAATTAAT 121  
Db 81 ATTTCACTGTGATGATATGTTGTCGAAAAAATAAGTCTTTGTTAAATTAAT 140  
OY 122 TGGTTGTGATCATTCTCTTTTCCCATTTGGAACATGATTAACCATCTCGAA 181  
Db 141 TGGTTGTGATCATTCTCTTTTCCCATTTGGAACATGATTAACCATCTCGAA 200  
OY 182 CTCGTAGAAAAACATCTGAAGAGCTAGTCTATGACATCTGACAGTGAATGGATGTT 241  
Db 201 CTGTGAAAAAATCTGTAAGAGCTAGTCTATGACATCTGACAGTGAATGGATGTT 260  
OY 242 CTCGAAACATTTCCACCAGACACCTGTTTCTATCCTGTTTAATTAATTAGTTGGTT 301  
Db 261 CTCGAAACATTTCCACCAGACACCTGTTTCTATCCTGTTTAATTAATTAGTTGGTT 320  
OY 302 CTCATCATGATTAACAACCCCTGCTCCATCTGTACATAAAGCTGTGACTTGAAGTT 361  
Db 321 CTCATCATGATTAACAACCCCTGCTCCATCTGTACATAAAGCTGTGACTTGAAGTT 380  
OY 362 TAGTC 366  
Db 381 TAGTC 385

RESULT 4  
LOCUS AA813266/c 599 bp mRNA EST 31-DEC-1998  
DEFINITION a144603.s1 Soares\_testis\_NHT Homo sapiens CDNA clone 1393204 3',  
mRNA sequence.  
ACCESSION AA813266  
NID 92883251  
VERSION AA813266.1 GI:2883251  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 599)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
On Apr 18, 1995 this sequence version replaced gi:775316.

JOURNAL  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/bdrip/image/image.html

Insert Length: 982 Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from AmerSham  
High quality sequence stop: 503.  
Location/Qualifiers  
1. .599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="1393204"  
/clone\_1lb="Soares\_testis\_NHT"  
/sex="male"

FEATURES  
source

/lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dT)  
 primer (5'  
 TGTACCAATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3').  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

202 a 98 c 131 g 168 t

Query Match 96.7%; Score 354; DB 39; Length 599;

Best Local Similarity 99.7%; Pred. No. 2.6e-71; Mismatches 0; Indels 1; Gaps 1;

Matches 365; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

2 CTTCTCTCCCTCCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTACAC 61  
 |||||||  
 454 CTTCTCTCCCTCCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTACAC 395  
 |||||||  
 62 ATTTCAGTGTATGATGTTGTC-AAAAAAAAAAAGTCTTTGTTAAATTAC 120  
 |||||||  
 394 ATTTCAGTGTATGATGTTGTC-AAAAAAAAAAAGTCTTTGTTAAATTAC 335  
 |||||||  
 121 TTGTTGTTGTAATTCATCTTTTCCCATTTGGAATGATTAATACCTCTCTGA 180  
 |||||||  
 334 TTGTTGTTGTAATTCATCTTTTCCCATTTGGAATGATTAATACCTCTCTGA 275  
 |||||||  
 181 ACTGTAGAAAAACATCTGAAAGAGTACTATCAGCATCTGACAGTGAATGGATGT 240  
 |||||||  
 274 ACTGTAGAAAAACATCTGAAAGAGTACTATCAGCATCTGACAGTGAATGGATGT 215  
 |||||||  
 241 TCTCAGAACCATTTACACCCAGACAGCTGTTTCTATCCCTGTTAAATTAATTTGGGT 300  
 |||||||  
 214 TCTCAGAACCATTTACACCCAGACAGCTGTTTCTATCCCTGTTAAATTAATTTGGGT 155  
 |||||||  
 301 TCTCTACATGATTAACAAACCCCTGCTCCATCTGTCTCATATAAAGTGTGACTGAACT 360  
 |||||||  
 154 TCTCTACATGATTAACAAACCCCTGCTCCATCTGTCTCATATAAAGTGTGACTGAACT 95  
 |||||||  
 361 TTAGTC 366  
 |||||||  
 94 TTAGTC 89

RESULT 5  
 AA437224 490 bp mRNA EST 30-MAY-1997  
 LOCUS AA437224.1  
 DEFINITION 3 similar to contains Alu repetitive element; mRNA sequence.  
 ACCESSION AA437224  
 NID 92142138  
 VERSION AA437224.1 GI:2142138  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 1 (bases 1 to 490)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Matra, M., Martin, J.,  
 Moore, B., Schellenberg, R., Stepien, M., Tan, F., Theisling, B.,  
 White, Y., Wylie, T., Waterston, R., and Wilson, R.  
 WashU-Merck EST Project 1997  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1405293.

Contact: Wilson RK  
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estevan@wustl.edu  
 This clone is available royalty-free through LINT; contact the  
 IMAGE Consortium (info@image.lit.gov) for further information.  
 Seq primer: -41ml3 fwd. Et from Amersham  
 High quality sequence stop: 363.  
 Location/Qualifiers  
 1. 490  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5978049"  
 /db\_xref="taxon:9606"  
 /map="6 p23-p22: 949D05; 2; 2q14.3-2q21.3"  
 /clone="IMAGE:757435"  
 /clone\_1lb="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - oligo(dT)  
 primer (5'  
 TGTACCAATCTGAGTGGAGCGCGCCCAATTTTCTTTTCTTTT 3').  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

173 a 73 c 101 g 143 t

Query Match 96.2%; Score 352; DB 33; Length 490;

Best Local Similarity 99.2%; Pred. No. 7.7e-71; Mismatches 0; Indels 3; Gaps 1;

Matches 365; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

2 CTTCTCTCCCTCCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTACAC 61  
 |||||||  
 453 CTTCTCTCCCTCCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTACAC 394  
 |||||||  
 62 ATTTCAGTGTATGATGTTGTC-AAAAAAAAAAAGTCTTTGTTAAATT 118  
 |||||||  
 393 ATTTCAGTGTATGATGTTGTC-AAAAAAAAAAAGTCTTTGTTAAATT 334  
 |||||||  
 119 ACTGTGTTGTAATTCATCTTTTCCCATTTGGAATGATTAATACCTCTCT 178  
 |||||||  
 333 ACTGTGTTGTAATTCATCTTTTCCCATTTGGAATGATTAATACCTCTCT 274  
 |||||||  
 179 GAACCTGTAGAAAAACATCTGAAAGAGTACTATCAGCATCTGACAGTGAATGGATG 238  
 |||||||  
 273 GAACCTGTAGAAAAACATCTGAAAGAGTACTATCAGCATCTGACAGTGAATGGATG 214  
 |||||||  
 239 GTTCTCAGAACCATTTACACCCAGACAGCTGTTTCTATCCCTGTTAAATTAATTTGG 298  
 |||||||  
 213 GTTCTCAGAACCATTTACACCCAGACAGCTGTTTCTATCCCTGTTAAATTAATTTGG 154  
 |||||||  
 299 GTTCTCAGAACCATTTACACCCAGACAGCTGTTTCTATCCCTGTTAAATTAATTTGG 358  
 |||||||  
 153 GTTCTCAGAACCATTTACACCCAGACAGCTGTTTCTATCCCTGTTAAATTAATTTGG 94  
 |||||||  
 359 GTTACTG 366  
 |||||||  
 93 GTTACTG 86

RESULT 6  
 A1685682 423 bp mRNA EST 27-MAY-1999  
 LOCUS A1685682.1  
 DEFINITION tt89h03.x1 nci\_cgap\_p128 Homo sapiens cDNA clone IMAGE:2248757 3',  
 mRNA sequence.  
 ACCESSION A1685682  
 NID 94896976



QY 100 AGTGTCTTTGTTAAATACCTGTTGTGAATCCATCTGCTTTTCCCATGGAC 159  
 Db 359 AGTGTCTTTGTTAAATACCTGTTGTGAATCCATCTGCTTTTCCCATGGAC 360  
 QY 160 TACTCATTAACCATCTCTGACCTGTAGAAAAACATCTGAAGCTAGCTATCAGCAT 219  
 Db 299 TACTCATTAACCATCTCTGACCTGTAGAAAAACATCTGAAGCTAGCTATCAGCAT 240  
 QY 220 CTGACAGGTAATGGATGGTCTCTCAGAACCATTTCCACCAGACGCTGTTCTATCCCT 279  
 Db 239 CTGACAGGTAATGGATGGTCTCTCAGAACCATTTCCACCAGACGCTGTTCTATCCCT 180  
 QY 280 GTTTAAATTAATTAATGTTGGTCTCTACATGATTAACAAACCTGCTCAATCTGCACA 339  
 Db 179 GTTTAAATTAATTAATGTTGGTCTCTACATGATTAACAAACCTGCTCAATCTGCACA 120  
 QY 340 TAAAGCTCTGACTGGAAGTTAGTC 366  
 Db 119 TAAAGCTCTGACTGGAAGTTAGTC 93  
 RESULT 8  
 LOCUS AA640928 411 bp mRNA EST 27-OCT-1997  
 DEFINITION n128d08.r1 NCI-CGAP\_P3 Homo sapiens CDNA clone IMAGE:1169295, mRNA  
 sequence.  
 ACCESSION AA640928  
 NID 92566178  
 VERSION AA640928.1 GI:2566178  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 411)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 30, 1996 this sequence version replaced gi:1530955.

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui,  
 M.D., Michael Emert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 CDNA Sequencing by: Genome Systems Inc., Greg Lennon, Ph.D.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 www.bio.llnl.gov/bbrp/image/image.html  
 Seg primer: -28m13 rev1 ET from Amersham  
 High quality sequence stop: 387.  
 Location/Qualifiers  
 1..411  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="1169295"  
 /clone\_lib="NCI-CGAP\_P3"  
 /sex="Male"  
 /dev\_stage="45 years old"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
 strand cDNA was primed with oligo(dT)17 on 50 ng of  
 DNase-treated, total cellular RNA obtained from  
 5,000-10,000 microdissected cells  
 histologically-determined to be fully malignant prostate  
 cancer cells. Double-stranded cDNA was ligated to EcoRI  
 adaptors, 5 cycles of PCR applied to the cDNA with an  
 adaptor-specific primer, and the resulting PCR product  
 subcloned into PAMP10 by the UDP-cloning method (Life

BASE COUNT 123 a 80 c 68 g 140 t  
 ORIGIN  
 Technologies). Average insert size is 600 bp. NOTE: Not  
 directionally cloned. This library was constructed by  
 David Kitzman.

Query Match 83.2%; Score 304.4; DB 36; Length 411;  
 Best Local Similarity 99.4%; Pred. No. 5e-60;  
 Matches 316; Conservative 0; Mismatches 1; Indels 1; Gaps 1.

QY 50 CAAGGATTACACATTTCACTGATGATATATGTTGTC-AAAAAAAAAAGTCTTT 108  
 Db 5 CAAGGATTACACATTTCACTGATGATATATGTTGCAAAAAAAGTCTTT 64  
 QY 109 GTTTAAATTAATTAATGTTGGTCTCTACATGATTAACAAACCTGCTCAATCTGCACA 168  
 Db 65 GTTTAAATTAATTAATGTTGGTCTCTACATGATTAACAAACCTGCTCAATCTGCACA 124  
 QY 169 ACCCATCTGTAAGTGTGTAAGAAACATCTGAAGCTAGTCTACAGATCTGACAGT 228  
 Db 125 ACCCATCTGTAAGTGTGTAAGAAACATCTGAAGCTAGTCTACAGATCTGACAGT 184  
 QY 229 GAATGATGATGTTCTCAGAACCATTTCCACCAGACGCTGTTCTATCCTGTTAATAA 288  
 Db 185 GAATGATGATGTTCTCAGAACCATTTCCACCAGACGCTGTTCTATCCTGTTAATAA 244  
 QY 289 ATTAGTTGGTCTCTACATGATTAACAAACCTGCTCAATCTGCACAATAAGTCT 348  
 Db 245 ATTAGTTGGTCTCTACATGATTAACAAACCTGCTCAATCTGCACAATAAGTCT 304  
 QY 349 GTGACTGGAAGTTAGTC 366  
 Db 305 GTGACTGGAAGTTAGTC 322

RESULT 9  
 LOCUS A1672753/c 383 bp mRNA EST 18-MAY-1999  
 DEFINITION w656d10.x1 Soares\_thymus\_NHEFH Homo sapiens CDNA clone  
 sequence.  
 ACCESSION A1672753  
 NID 94852484  
 VERSION A1672753.1 GI:4852484  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 383)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3137861.

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 This clone is available royalty-free through LLNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seg primer: -40up from Gibco  
 High quality sequence stop: 362.  
 Location/Qualifiers  
 1..383  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="952F05; 14; 14q23.1-14q24.1; MMU16C3-C4 region;  
 MMU16C3-C4 region"  
 /clone\_image="1169295"  
 /clone\_lib="Soares\_thymus\_NHEFH"  
 /dev\_stage="fetal"



/lab\_host="DH10B (phage-resistant)"  
 /note="Organ: thymus, pooled; Vector: pT73D-Pac  
 (Pharmacia) with a modified polylinker; Site 1: Not I;  
 Site 2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5',  
 TGTTACCAATCTGAGAGTGGAGCGCCGACAGCTTTTCTTTTCTTTT 3'],  
 double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization. Library  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 134 a 63 c 75 g 111 t  
 ORIGIN

Query Match 81.7% Score 299; DB 50; Length 383;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-59;

Matches 299; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 CTGTGATGATATATGTTGCAAAAAAAGTGTCTTTTAAATTAATCTGGTTT 127  
 |||||||  
 DB 383 CTGTGATGATATATGTTGCAAAAAAAGTGTCTTTTAAATTAATCTGGTTT 324  
 |||||||  
 QY 128 GTGAATCCATCTGCTTTTCCCATTTGGAATAGTCAATTAACCCATCTGAACGTGA 187  
 |||||||  
 DB 323 GTGAATCCATCTGCTTTTCCCATTTGGAATAGTCAATTAACCCATCTGAACGTGA 264  
 |||||||  
 QY 188 GAATAACATCTGAAGAGTACTATATGATCTGACAGTGCATTTGGATGGTCTCGA 247  
 |||||||  
 DB 263 GAATAACATCTGAAGAGTACTATATGATCTGACAGTGCATTTGGATGGTCTCGA 204  
 |||||||  
 QY 248 ACCATTTCACCCAGACAGCCGTCTTCTATCTGTTTAATAATTAGTGGTCTCTAC 307  
 |||||||  
 DB 203 ACCATTTCACCCAGACAGCCGTCTTCTATCTGTTTAATAATTAGTGGTCTCTAC 144  
 |||||||  
 QY 308 ATGCATTAACAAACCCGTCTCAATCTGTGACATAAAAGTGTGACTGAAGTTAGTC 366  
 |||||||  
 DB 143 ATGCATTAACAAACCCGTCTCAATCTGTGACATAAAAGTGTGACTGAAGTTAGTC 85  
 |||||||

RESULT 10  
 AA533772/c 445 bp mRNA EST 21-AUG-1997  
 LOCUS n193c05.s1 NCI\_CGAP\_Pr11 Homo sapiens cDNA clone IMAGE:1000040,  
 DEFINITION mRNA sequence.

ACCESSION AA533772  
 NID 92277788  
 VERSION AA533772.1 GI:2277788  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 445)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE Tumor Gene Index  
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 COMMENT Unpublished (1997)  
 On May 8, 1995 this sequence version replaced gi:800232.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Rodrigo F.

Chuaquil, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center

found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.lnl.gov/bdrp/image/image.html

Insert Length: 613 Std Error: 0.00

Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 317.  
 Location/Qualifiers

FEATURES  
 source

1.445  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1000040"  
 /clone\_1ib="NCI\_CGAP\_Pr11"  
 /sex="male"

/tissue\_type="normal prostatic epithelial cells"

/lab\_host="DH10B"  
 /note="Organ: prostate; Vector: PAMP10; mRNA made from  
 normal prostatic epithelial cells, cDNA made by oligo-dT  
 priming. Non-directionally cloned. Size-selected on  
 agarose gel, average insert size 600 bp. Library made by  
 D. Kitzman, NIH."

BASE COUNT 148 a 69 c 87 g 141 t  
 ORIGIN

Query Match 80.4% Score 294.2; DB 34; Length 445;  
 Best Local Similarity 98.5%; Pred. No. 1e-57;

Matches 318; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 45 ATTTGCAAGGATTTACATTTCACTGTGATGATATGTTGCAAAAAAAGTGTG- 104  
 |||||||  
 DB 421 AATTGCAAGGATTTACATTTTAC-GGATGATATATGTTGCAAAAAAAGTGTG- 363  
 |||||||  
 QY 104 TCTTTGTTTAAATTAATCTGTTGTTGTAATCATCTGCTTTTCCCATTTGAATAGT 163  
 |||||||  
 DB 362 TCTTTGTTTAAATTAATCTGTTGTTGTAATCATCTGCTTTTCCCATTTGAATAGT 303  
 |||||||  
 QY 164 CATTAACCCATCTGTGACGTGTAAGAAACATCTGTAAGAGCTAGTATCAGCATCTGA 223  
 |||||||  
 DB 302 CATTAACCCATCTGTGACGTGTAAGAAACATCTGTAAGAGCTAGTATCAGCATCTGA 243  
 |||||||  
 QY 224 CAGGTGAATGAGATGTTCTCAGAACCATTTTACCCAGACGCTGTTTATCTGTTT 283  
 |||||||  
 DB 242 CAGGTGAATGAGATGTTCTCAGAACCATTTTACCCAGACGCTGTTTATCTGTTT 183  
 |||||||  
 QY 284 AATTAATTAATGTTGGTCTCTACATCATTAACAAACCCGTCTCAATCTGTGCATATAA 343  
 |||||||  
 DB 182 AATTAATTAATGTTGGTCTCTACATCATTAACAAACCCGTCTCAATCTGTGCATATAA 123  
 |||||||  
 QY 344 AGTCTGTGACTGAAGTTAGTC 366  
 |||||||  
 DB 122 AGTCTGTGACTGAAGTTAGTC 100  
 |||||||

RESULT 11  
 AA531606/c 423 bp mRNA EST 20-AUG-1997  
 LOCUS n165604.s1 NCI\_CGAP\_Pr10 Homo sapiens cDNA clone IMAGE:997374, mRNA  
 DEFINITION sequence.

ACCESSION AA531606  
 NID 92274312  
 VERSION AA531606.1 GI:2274312  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 423)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:139565.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550

Email: Robert\_Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuaquil,



ACCESSION AA650104  
 MID 92577432  
 VERSION AA650104.1 GI:2577432  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 339)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 1, 1995 this sequence version replaced.

COMMENT  
 JOURNAL  
 CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquai,  
 M.D., Michael Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)

Insert Length: 1362 Std Error: 0.00  
 Seq primer: -40m13 fwd. ET from Amersham  
 High quality sequence stop: 291.

## FEATURES

source  
 1.339  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="17q21: between D17S1321 and D17S1325"  
 /clone="IMAGE:119117"  
 /clone\_lib="NCI-CGAP\_Pr3"  
 /sex="Male"  
 /dev\_stage="45 years old"  
 /lab\_host="DH10B"  
 /note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
 strand cDNA was primed with 0.190(47)17 on 50 ng of  
 DNase-treated, total cellular RNA obtained from  
 5,000-10,000 microdissected cells  
 histologically-determined to be fully malignant prostate  
 cancer cells. Double-stranded cDNA was ligated to EcoRI  
 adaptors, 5 cycles of PCR applied to the cDNA with an  
 adaptor-specific primer, and the resulting PCR product  
 subcloned into PAMP10 by the UDG-cloning method (Life  
 Technologies). Average insert size is 600 bp. NOTE: Not  
 directionally cloned. This library was constructed by  
 David Kitzman."

BASE COUNT 94 a 72 c 57 g 116 t

ORIGIN

Query Match 77.7% Score 284.2; DB 36; Length 339;  
 Best Local Similarity 99.0% Pred. No. 2e-55;  
 Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
 Oy 2 CTTCTCTCCCTCCCTCGAATTAATCTTCACTTGCAATTTGCAAGATACAC 61  
 |||||  
 Db 51 CTCGTCTCCCTCCCTCGAATTAATCTTCACTTGCAATTTGCAAGATACAC 110  
 Oy 62 ATTCACTGATGATATATGTTGTCAGAAAAAGTCTTTGTTAAATTAAT 121  
 |||||  
 Db 111 ATTCACTGATGATATATGTTGTCAGAAAAAGTCTTTGTTAAATTAAT 170  
 Oy 122 TGGTTGTGAATCATGCTTTTCCCATGGAATGATTAACCATCTCGAA 181  
 |||||  
 Db 171 TGGTTGTGAATCATGCTTTTCCCATGGAATGATTAACCATCTCGAA 230  
 Oy 182 CTGTAGAAAAACATCTGAAGAGCTAGTCTATCAGCATCTGACAGTGAATGATGTT 241

Db 231 |||||  
 CTCGTAGAAAAACATCTGAAGAGCTAGTCTATCAGCATCTGACAGTGAATGATGTT 290  
 Oy 242 |||||  
 CTCGAGAACCATTTCCACAGACCTGTTCTATCTCTTAAAT 290  
 Db 291 CTCGAGAACCATTTCCACAGACCTGTTCTATCTCTTAAAT 339

RESULT 14  
 AI669511/c 377 bp mRNA EST 14-MAY-1999  
 LOCUS WB88f07.x1 NCI-CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2312773 3',  
 DEFINITION mRNA sequence.  
 ACCESSION AI669511  
 NID 94834285  
 VERSION AI669511.1 GI:4834285  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 1 (bases 1 to 377)  
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Dec 20, 1995 this sequence version replaced gi:1130922.

## FEATURES

source  
 Seq primer: -40UP from Gibco.  
 Location/Qualifiers  
 1.377  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2312773"  
 /clone\_lib="NCI-CGAP\_Pr28"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI-CGAP\_Pr22 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clones  
 96508-966759, 1101192-1101959, and 1217928-1220515).  
 Subtraction by Bento Soares and M. Fatima Bonaldi."  
 www-bio.llnl.gov/dbp/image/image.html

## FEATURES

source  
 Seq primer: -40UP from Gibco.  
 Location/Qualifiers  
 1.377  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2312773"  
 /clone\_lib="NCI-CGAP\_Pr28"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI-CGAP\_Pr22 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clones  
 96508-966759, 1101192-1101959, and 1217928-1220515).  
 Subtraction by Bento Soares and M. Fatima Bonaldi."  
 www-bio.llnl.gov/dbp/image/image.html

## BASE COUNT

132 a 61 c 74 g 110 t

ORIGIN

Query Match 76.8% Score 281; DB 50; Length 377;  
 Best Local Similarity 99.7% Pred. No. 1e-54;  
 Matches 292; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 Oy 75 GATATATGTTGTC-AAAAAAGTCTTTGTTAAATTAATCTGTTGTAAT 133  
 |||||  
 Db 377 GATATATGTTGTCAGAAAAAAGTCTTTGTTAAATTAATCTGTTGTAAT 318  
 Oy 134 CCAATCTGCTTTTCCCATGGAATGATTAACCATCTCTGAACTGAGAAAA 193  
 |||||

Db 317 CCATCTTGCTTTTCCCATTTGGAAGTATGATTAACCATCTCTGAAGTGTAGAAAA 258  
 QY 194 CATCTGAGAGTAGTATCATCAGCATCTGACAGGTGAATGGATGGTTCAGAACCAT 253  
 Db 257 CATCTGAGAGTAGTATCATCAGCATCTGACAGGTGAATGGATGGTTCAGAACCAT 198  
 QY 254 TCACCCAGACAGCGCTGTTCTATCTGTTTAATTAATTAAGTTGGGTTCTCTACATGCAT 313  
 Db 197 TCACCCAGACAGCGCTGTTCTATCTGTTTAATTAATTAAGTTGGGTTCTCTACATGCAT 198  
 QY 314 ACAAACCCCTGCTCAATCTGTCACATAAAGTCTGTAAGTTAGTCTC 366  
 Db 137 ACAAACCCCTGCTCAATCTGTCACATAAAGTCTGTAAGTTAGTCTC 85

RESULT 15  
 W47380/c 374 bp mRNA EST 23-MAY-1996  
 LOCUS 2c39n11.s1 Soares\_senescent\_fibroblasts\_NBHSF Homo sapiens cDNA  
 DEFINITION clone IMAGE:324741 3' similar to contains Alu repetitive element.;  
 mRNA sequence.  
 ACCESSION W47380  
 NID G1332019  
 VERSION W47380.1 GI:1332019  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 374)  
 AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,  
 Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,  
 Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,  
 Trevisakis,E., Waterston,R., Williamson,A., Wohlmann,P. and  
 Wilson,R.  
 / The WashU-Merck EST Project  
 JOURNAL Unpublished (1995)  
 COMMENT  
 Contact: Wilson RK  
 Washington University School of Medicine  
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: est@wustl.edu  
 This clone is available royalty-free through LNL: contact the  
 IMAGE Consortium (info@image.lnl.gov) for further information.  
 Seq primer: mob.REGA+FT  
 High quality sequence stop: 360.  
 Location/Qualifiers  
 1. 374  
 /organism="Homo sapiens"  
 /db\_xref="GDB:1256253"  
 /db\_xref="taxon:9606"  
 /clone\_image="324741"  
 /clone\_lib="Soares\_senescent\_fibroblasts\_NBHSF"  
 /issue\_type="senescent\_fibroblast"  
 /lab\_host="DH10B (ampicillin resistant)"  
 /note="Vector: pTR3D (Pharmacia) with a modified  
 polylinker V-type phagemid; Site\_1: Not I; Site\_2: Eco  
 RI; 1st strand cDNA was primed with a Not I - oligo(dT)  
 primer [5'  
 TGTTCACATCTGAGTGGAGCGGCCGATTTTTTTTTTTTTTTTTTTT 3']  
 double-stranded cDNA was size selected, ligated to Eco RI  
 adapters (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of a modified pTR3D vector  
 (Pharmacia). Library went through one round of  
 normalization to a Cot = 5. Library constructed by Bento  
 Soares and M.Fatima Bonaldo."

BASE COUNT 131 a 59 c 74 g 110 t  
 ORIGIN

Query Match<sub>J</sub>

76.5%; Score 280; DB 26; Length 374;

Best Local Similarity 99.7%; Pred. No. 1.8e-54;  
 Matches 291; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
 QY 76 TATATGTTTGC-AAAAAAAAAAGTCTCTTGTGTTAAATTAATCTGTTGGAAATC 134  
 Db 374 TATATGTTTGCAAAAAAAAAAGTCTCTTGTGTTAAATTAATCTGTTGGAAATC 315  
 QY 135 CATCTGCTTTTCCCATTTGGAAGTATGATTAACCATCTCTGAAGTGTAGAAAAAC 194  
 Db 314 CATCTGCTTTTCCCATTTGGAAGTATGATTAACCATCTCTGAAGTGTAGAAAAAC 255  
 QY 195 ATCTGAAGAGTAGTCTATCAGCATCTGACAGGTGAATGGATGGTTCACAGAACATTT 254  
 Db 254 ATCTGAAGAGTAGTCTATCAGCATCTGACAGGTGAATGGATGGTTCACAGAACATTT 195  
 QY 255 CACCCAGACAGCGCTGTTCTATCTGTTTAATTAATTAAGTTGGGTTCTCTACATGCATA 314  
 Db 194 CACCCAGACAGCGCTGTTCTATCTGTTTAATTAATTAAGTTGGGTTCTCTACATGCATA 135  
 QY 315 ACAAACCCCTGCTCAATCTGTCACATAAAGTCTGTAAGTTAGTCTC 366  
 Db 134 ACAAACCCCTGCTCAATCTGTCACATAAAGTCTGTAAGTTAGTCTC 83

Search completed: September 28, 1999, 11:29:30  
 Job time: 1851 sec

Wed Sep 29 14:27:11 1999

us-09-030-606-115.rst

---

Page 11

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compaq Inc.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:23:15; Search time 2910.9 Seconds  
(without alignments)  
399.874 Million cell updates/sec

Title: US-09-030-606-115

Perfect score: 366

Sequence: 1 GCTCTTCTCTCTCCCTCTC.....CTGTGCTTGAAGTTAGTC 366

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_om:\*  
4: gb\_ov:\*  
5: gb\_pac:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl3:\*  
10: gb\_pr2:\*  
11: gb\_pr3:\*  
12: gb\_ro:\*  
13: gb\_st:\*  
14: gb\_sts:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: gb\_fun:\*  
19: em\_hgc:\*  
20: em\_hum1:\*  
21: em\_hum2:\*  
22: em\_in:\*  
23: em\_om:\*  
24: em\_or:\*  
25: em\_ov:\*  
26: em\_pat:\*  
27: em\_ph:\*  
28: em\_pl:\*  
29: em\_ro:\*  
30: em\_sts:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: gb\_hgc1:\*  
35: gb\_hgc2:\*  
36: gb\_in1:\*  
37: gb\_in2:\*  
38: em\_ba1:\*  
39: em\_ba2:\*  
40: em\_hum3:\*  
41: em\_hum4:\*  
42: gb\_pre4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	280	76.5	374	14	G37241	G37241 SHGC-57295

2	45	12.3	3145	12	MMU88542	U88542 Mus musculus
C 3	43.6	11.9	93646	7	ATF18F4	AL021637 Arabidops
C 4	41.8	11.4	119707	35	AC007180	AC007180 Drosophila
C 5	41.8	11.4	77685	37	AC005444	AC005444 Drosophila
C 6	40.4	11.0	262336	34	AC006786	AC006786 Caenorhab
C 7	40.4	11.0	207370	34	AC006798	AC006798 Caenorhab
C 8	40.4	11.0	200509	34	AC006910	AC006910 Caenorhab
C 9	40.4	11.0	40479	36	CELK12C11	AF043701 Caenorhab
C 10	39.8	10.9	104386	9	D87009	D87009 Human (Lamb
C 11	39.6	10.8	39565	36	CEK08E3	281568 Caenorhabd
C 12	38.8	10.6	145920	34	AC006712	AC006712 Caenorhab
C 13	38.8	10.6	299919	34	AC006712	AC006712 Caenorhab
C 14	38.8	10.6	110000	34	CEY11B2.0	298857 Caenorhabd
C 15	38.8	10.6	110000	34	CEY11B2.1	Continuation (2 of
C 16	38.8	10.6	110000	34	CEY11B2.2	Continuation (3 of
C 17	38.8	10.6	28678	36	CEL104D1	AF067817 Caenorhab
C 18	38.6	10.5	28992	7	SC9375	Z47071 S.cerevisia
C 19	38.4	10.5	177411	11	AC005696	AC005696 Homo sapi
C 20	38.2	10.4	97451	8	F19K23	AC000375 Sequence
C 21	38.2	10.4	253176	34	PFMAL13P2	AL049185 Plasmodiu
C 22	38	10.4	13776	1	STARRSC	L11998 Staphylococ
C 23	38	10.4	46445	2	AF051917	AF051917 Staphyloc
C 24	38	10.4	288	14	DM163A55	231900 D. melanoga
C 25	37.8	10.4	34351	36	CELI14G12	U41268 Caenorhabd
C 26	37.8	10.3	209365	34	AC006719	AC006719 Caenorhab
C 27	37.8	10.3	22846	36	CELE04A4	AF038611 Caenorhab
C 28	37.4	10.2	55136	9	HUMVITDBP	L10641 Human vitam
C 29	37.4	10.2	9985	37	PCU4145	U43145 Plasmodium
C 30	37.2	10.2	37321	11	AC005498	AC005498 Homo sapi
C 31	37	10.1	76183	8	ATAC006570	AC006570 Arabidops
C 32	37	10.1	156300	11	AC005919	AC005919 Homo sapi
C 33	37	10.1	43410	36	CEC01G6	235595 Caenorhabd
C 34	37	10.1	4200	37	AR145052	AR145052 Plasmodiu
C 35	36.6	10.0	7218	5	I66494	I66494 Sequence 14
C 36	36.6	10.0	2157	36	TEMTMC	D12772 Tetrahymena
C 37	36.6	10.0	153206	42	AC006356	AC006356 Homo sapi
C 38	36.4	9.9	139444	9	HS380C13	AL022161 Human DNA
C 39	36.4	9.9	132090	12	AC005402	AC005402 Mus muscu
C 40	36.4	9.9	45348	34	AC006624	AC006624 Caenorhab
C 41	36.4	9.9	312267	34	AC006785	AC006785 Caenorhab
C 42	36.4	9.9	292390	34	AC006900	AC006900 Caenorhab
C 43	36.4	9.9	110000	34	CEY51A2.2	Continuation (3 of
C 44	36.4	9.9	139259	36	CEY51A2D	AL021497 Caenorhab
C 45	36.4	9.9	1528	36	TETMACAL	D12774 Tetrahymena

## ALIGNMENTS

RESULT 1  
G37241/c 374 bp DNA STS 30-MAR-1998  
LOCUS SHGC-57295 Human Homo sapiens STS genomic, sequence tagged site.  
DEFINITION G37241  
ACCESSION G2996892  
NID G37241.1 GI:2996892  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 374)  
Myers,R.M.  
Human STS (1997)  
Unpublished (1997)  
COMMENT  
Contact: Richard M. Myers  
Stanford Human Genome Center (SHGC)  
Stanford University School of Medicine  
Department of Genetics, W-344, Stanford, CA 94305, USA  
Tel: 4157259687  
Fax: 4157259689  
Email: myers@shgc.stanford.edu

Primer A: TGTGACAGATTGGAGCAGG  
Primer B: TCGTTTCCCATTTGGAAC  
STS size: 200  
PCR Profile:

Initial incubation: 95 degrees C for 10 minutes

Denaturation: 94 degrees C for 30 seconds  
Annealing: 60 degrees C for 30 seconds  
Polymerization: 72 degrees C for 23 seconds  
PCR Cycles: 30  
Thermal Cycler: Perkin Elmer 9700

Protocol:  
Template: 25 ng  
Primer: each 1 uM  
dNTPs: each 200 uM  
AmpliTaq Gold Polymerase: 0.07 units/uL  
Total Vol: 5 uL

Buffer:  
MgCl2: 2.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 8.3

FEATURES  
source Prepared with primer pairs derived from W47380 -- Unigene.  
Location/Qualifiers  
1..374  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="8"  
/clone\_11b="Human"

STS  
primer\_bind 110..129  
primer\_bind complement(290..309)

BASE COUNT 131 a 59 c 74 g 110 t  
ORIGIN

Query Match 76.5%; Score 280; DB 14; Length 374;  
Best Local Similarity 99.7%; Pred. No. 6.1e-53;  
Matches 291; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

OY 76 TATATTGTTGTC-AAAAAAAAAAGTGTCTTTTAAATTAAGTTGTTGTAATC 134  
DB 374 TATATTGTTGTCAAAAAAAAAAGTGTCTTTTAAATTAAGTTGTTGTAATC 315  
OY 135 CATCTTGTCTTCCCATGTGAGTATGATTAACCCATCTCTGAACTGTAGAAAAC 194  
DB 314 CATCTTGTCTTCCCATGTGAGTATGATTAACCCATCTCTGAACTGTAGAAAAC 255  
OY 195 ATCTGAAGAGTATCTATCAGCATCTGACAGTGAATGATGTTTCAGAACATT 254  
DB 254 ATCTGAAGAGTATCTATCAGCATCTGACAGTGAATGATGTTTCAGAACATT 195  
OY 255 CACCCAGACAGCCTGTTCTATCTGTTAAATAATTAAGTTGGGTTCTCTACATGCATA 314  
DB 194 CACCCAGACAGCCTGTTCTATCTGTTAAATAATTAAGTTGGGTTCTCTACATGCATA 135  
OY 315 ACAAAACCTGCTCAATCTGTACATATAAAGTGTGATTTGAAGTTTATGTC 366  
DB 134 ACAAAACCTGCTCAATCTGTACATATAAAGTGTGATTTGAAGTTTATGTC 83

RESULT 2  
LOCUS MW08542 3145 bp mRNA ROD 20-MAY-1997  
DEFINITION Mus musculus homeobox protein Nkx3.1 mRNA, complete cds.  
ACCESSION U88542  
VERSION 92105349  
KEYWORDS U88542.1 GI:2105349  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE  
AUTHORS Eukaryota: Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 3145)  
Scialvino, P.J., Abrams, E.W., Yang, L., Austenberg, L.P., Shen, M.M.  
and Abate-Shen, C.  
TITLE Tissue-specific expression of murine Nkx3.1 in the male urogenital  
system  
JOURNAL Dev. Dyn. 209 (1), 127-138 (1997)  
MEDLINE 97287410

REFERENCE  
AUTHORS Scialvino, P.J., Austenberg, L.P. and Abate-Shen, C.  
TITLE Direct Submission  
JOURNAL Submitted (04-FEB-1997) CABM, UNDNJ-RWJMS, 679 Hoes Lane,  
Piscataway, NJ 00854, USA

FEATURES  
source Location/Qualifiers  
1..3145  
/organism="Mus musculus"  
/strain="Swiss-Webster/NIH"  
/db\_xref="taxon:10090"  
14..727  
/codon\_start=1  
/product="homeobox protein Nkx3.1"  
/protein\_id="AAB58025.1"  
/db\_xref="PID:g2105350"  
/db\_xref="GI:2105350"  
/translation="MLRVAPREPRVAGGRSPMAAPPTOSKRRLTSFLIODIDIRRAE  
RHGGHSGNPDSPDRDASAPDDKAGRGVAPEDPSIRSHAPETPEESPAHET  
YLDGHNPGDLSAPDYTKOPKORSAASHOVIELEKFSHOKTSLSPERAHAK  
NKLRTQVAKIMFONRKYTKRKQSLSDLVLEKNSPLSLPALKDLSLPSTLSVYT  
SYPPYPLVLCISWHPSEW"  
386..565  
/note="encodes the homeobox"

CDS

misc\_feature  
BASE COUNT 769 a 760 c 794 g 822 t  
ORIGIN

Query Match 12.3%; Score 45; DB 12; Length 3145;  
Best Local Similarity 63.9%; Pred. No. 0.41;  
Matches 85; Conservative 0; Mismatches 45; Indels 3; Gaps 1;

OY 62 ATTACCTGATGATATATTTGTGTGCAAAAAAAAAAAGTCTTTGTTAAATTAATCT 121  
DB 2777 ATTATAGTGTGACAGTGTGTGCAAAAGGAAAGGCTGTGTTAAATTAATGT 2836  
OY 122 TGG--TTTGAATTCATCTTGTCTTCCCAATGGAATGATCAATCAATCATCTCT 178  
DB 2837 TGGGAAGTCTGACTCAATTTTGTCTTCTCCATCTGGAATCAATCAATCAATCT 2896  
OY 179 GAACGTGTAGAAA 191  
DB 2897 GAACGTGTAGAAA 2909

RESULT 3  
LOCUS AF18F4/c 93646 bp DNA PLN 10-FEB-1999  
DEFINITION Arabidopsis thaliana DNA chromosome 4, BAC clone F18F4 (ESSAII  
project).  
ACCESSION AF18F4  
VERSION 92827644  
KEYWORDS AF18F4.1 GI:2827644  
SOURCE thale cress.  
ORGANISM Arabidopsis thaliana  
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;  
Rosidae; Caprariales; Brassicaceae; Arabidopsis.  
REFERENCE  
AUTHORS Bevan, M., Murphy, G., Drost, L., Hall, C., Hudson, S., Ridley, P.,  
Bancroft, I., Mewes, H.W., Mayer, K. and Scheller, C.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 93646)  
AUTHORS EU Arabidopsis sequencing project.



TITLE Direct Submission  
JOURNAL Submitted (03-FEB-1998) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk

FEATURES  
source  
1. .93646  
/organism="Arabidopsis thaliana"  
/variety="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="4"  
138. .3270  
/gene="F18F4.10"  
complement(join(138. .251,322. .470,2694. .2855,3219. .3270))  
/gene="F18F4.10"  
/note="Protein sequence is in conflict with the conceptual translation."  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CA16599.1"  
/db\_xref="PID:g1248720"  
/db\_xref="PID:g2827645"  
/db\_xref="GI:2827645"  
/db\_xref="SPTREMBL:O49418"  
/translation="MAKLCSYLFICMFYLSGFVSASAKOLKTCSTVIRKIGHPCDIE  
SCINCEFRVNTGFAFGDQVSGDCTCPNPVVALDELVKIQLYDKQLYTIPISY  
KLETSVRLIKQEGFNGFSPSNAYIDVNMILDVKNLQDSAGCGTMLKLFACS"  
complement(141. .251)  
/gene="F18F4.10"  
/number=1  
complement(252. .321)  
/gene="F18F4.10"  
/number=1  
complement(322. .470)  
/gene="F18F4.10"  
/number=2  
complement(471. .2693)  
/gene="F18F4.10"  
/number=2  
complement(2694. .2855)  
/gene="F18F4.10"  
/number=3  
complement(2856. .3218)  
/gene="F18F4.10"  
/number=3  
complement(3219. .3270)  
/gene="F18F4.10"  
/number=4  
4337. .9232  
/gene="F18F4.20"  
complement(4337. .4424)  
/gene="F18F4.20"  
/number=1  
complement(join(4337. .4424,4499. .4645,4739. .5174,  
5945. .6138,6283. .6882,8391. .8639,8742. .9232))  
/gene="F18F4.20"  
/note="similarity to downy mildew resistance protein RPP5,  
Arabidopsis thaliana, PATCH:G2109275"  
/codon\_start=1  
/product="downy mildew resistance-like protein"  
/protein\_id="CA16600.1"  
/db\_xref="PID:g1248721"  
/db\_xref="PID:g2827646"  
/db\_xref="GI:2827646"  
/db\_xref="SPTREMBL:O49419"  
/translation="MPAFRRPRPQIYISFGNDLRKGFVSHVVKALKDARVAVFDN  
DHQWRGRGHHDDHLFVRINSLKALVIFSDYAESQCLNELTTHIRYAEGLK  
MVTPIFYKVIIEVNNLEEGFGKCPDEMVTQORHQLTHIYGLIRSAKRPST  
GYISNDLVEATITIGIKKIPYISAKQITIGEEVSAELASVVAALCFITISVFP  
DLGFIISPRWYLAVMLEFVVRHLSIQQRPMDDPPSPPPPIPESRPRLTPPVILT  
RPPPLPYARLPQSLPPRPPOVFSFGKELHGFVSHVVAALRIGVAVFIDS  
NEKGRDLQNLFRKRIENSKMALVIFSDRFESDMCLNELVKIDCVKEGLVYIPVFY

RVNFDVKNKGFSGFIEIVQROSPKEEPAERWNSVKSISKTGTFSEVHRIDS  
YLVDAIVRDKROLVPYPTKEKLEPIETEIFPALYAGCNFIAPLIFNDTSFYTPQ  
WVFVFNHGESDSGDFIRHVERALNDEGFNFIDSDERGRGMHEIFPAINSVAL  
VTFSDRYTSEICLIEHARIVYDRRREGKYLVPVTRYSDVNNFNGRFGSFLFT  
TIQGFEDHPFAHMMNWNVEICTDGFYSADYSNDTSIVEIVRAIKRRLQKGLPR  
RVODEKLPSPDLMFMPKOPNAQGDVFSILOWLVAVLSLVIRRTFRLLIS"  
complement(4425. .4498)  
/gene="F18F4.20"  
/number=1  
complement(4499. .4645)  
/gene="F18F4.20"  
/number=2  
complement(4646. .4738)  
/gene="F18F4.20"  
/number=2  
complement(4739. .5174)  
/gene="F18F4.20"  
/number=3  
complement(5175. .5944)  
/gene="F18F4.20"  
/number=3  
complement(5945. .6138)  
/gene="F18F4.20"  
/number=4  
complement(6139. .6282)  
/gene="F18F4.20"  
complement(6283. .6882)  
/gene="F18F4.20"  
/number=5  
complement(6883. .8390)  
/gene="F18F4.20"  
/number=5  
complement(8391. .8639)  
/gene="F18F4.20"  
/number=6  
complement(8640. .8741)  
/gene="F18F4.20"  
/number=6  
complement(8742. .9232)  
/gene="F18F4.20"  
/number=7  
12070. .13365  
/gene="F18F4.30"  
/number=1  
12070. .13365  
/gene="F18F4.30"  
/number=1  
12070. .13365  
/gene="F18F4.30"  
/note="similarity to W20DMY30S"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CA16601.1"  
/db\_xref="PID:g1248722"  
/db\_xref="PID:g2827647"  
/db\_xref="GI:2827647"  
/db\_xref="SPTREMBL:O49420"  
/translation="MAFSFDLAVIFQREMRKRKSHVYIRRRRRDMCKSHBPMYIP  
FDLVLEILRLPAKSLMRKSVSKLWSSLCISRTFTNRLRYPSFQRIYVLTFLDN  
SLORKSLSSSSPGSDISYSSFYVDRDLTPSMKGYLSHVLRGLMCFKPEKSVK  
IYNTTROIIVLPDIEESNIIAEDHKNKIMYRIGHDPVGDQYKVCYIARNDERGE  
LRYLSEHNVFTLGGCSSGMRKIPCPSPHLLITOLISNGRHHIYAWQKDPMLYT  
FDFSSSEIISIIQAPEDIRWFKSNPIETGVALLNLSLDRCTMLWAVEVEKNMW  
SEKTLVHPDSQMDIYKSTLRVAGTTRNNEVILVPNINIRYTLTGEIVYEPONTLLYI  
14707. .15942  
/gene="F18F4.40"  
/number=1  
14707. .15942  
/gene="F18F4.40"  
14707. .15942  
/gene="F18F4.40"  
/note="similarity to W15DMY30S"



FEATURES

76032 76131: gap of unknown length

76132 96831: contig of 20700 bp in length

96832 96911: gap of unknown length

96912 119707: contig of 22796 bp in length.

Location/Qualifiers

1. 119707

/organism="Drosophila melanogaster"

/strain="y2: cn bw sp"

/db.xref="taxon:7227"

/clone="BACR01G10 (D539) RPCI-98 01.G.10"

/chromosome="2"

/clone\_1lb="RPCI-98 (Roswell Park Cancer Institute Drosophila melanogaster BAC library, partial ECORI in pBac3.6"

/map="36D-36D"

BASE COUNT 34438 a 24821 c 24376 g 34471 t 1601 others

ORIGIN

Query Match 11.4% Score 41.8; DB 35; Length 119707;

Best Local Similarity 55.0%; Pred. No. 1.1;

Matches 82: Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 1 GCTCTTCTCTCCCTCCTGAAATTAATCTTTCACATTCGCAATTCGACAGATTACA 60

DB 88106 GATAATTCACGATGATGATCTTAAGTATGATTAATGATTAATGATTAAGAAATAAA 88047

OY 61 CATTTCACTGATGATATATGTGTGCAAAAAAAGTCTTCTTTAAATATAC 120

DB 88046 CACCTTGCATTAATTAATGATGTTTATTTACAAATAAACAATCTTTAATAAGTAC 87987

OY 121 TTGGTTGTGATCCATCTTCTTTTCC 149

DB 87986 TTACTTTATACCCAGTTTCTTGCCCC 87958

RESULT 5

AC005444/c 77685 bp DNA INV 31-OCT-1998

LOCUS Drosophila melanogaster, chromosome 2R, region 36D3-36D3, Pl clone

DEFINITION DS06379, complete sequence.

AC005444

NID 93818341

VERSION AC005444.1 GI:3818341

KEYWORDS HTG.

SOURCE fruit fly

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 77685)

Celniker,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazey,R.G.,

Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,

Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,

Hoskins,R.A., Houston,K.A., Hummst,S.R., Karra,K., Kearney,L.,

Kim,E., Lee,B., Lewis,S., Li,P., Lomoton,M.A., Mazda,P.,

Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,

Pfeiffer,B., Poon,L., PUNCH,E., Sequella,A., Sethi,R., Sht,E.,

Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,

Zleran,L.L. and Rubin,G.M.

Sequencing of Drosophila chromosome 2R, region 36D3-36D3

Unpublished (1998)

2 (bases 1 to 77685)

Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,

Agbayan,A., Arcaina,T.T., Baxter,E., Blazey,R.G., Chavez,C.,

Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,

Hummst,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,

Lomoton,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,

Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., PUNCH,E.,

Sht,E., Twomey,B., Wan,K.H., Whiteley,K.R., Yee,A., Zhang,R.,

Zleran,L.L. and Kimmel,B.E.

Direct Submission

Submitted (14-AUG-1998) Drosophila Genome Center, Lawrence Berkeley

COMMENT

Laboratory, MS 64-121, Berkeley, CA 94720, USA

On Oct 31, 1998 this sequence version replaced g1:3805723.

Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121

Berkeley, CA 94720

For further information about this sequence, including its location

and relationship to other sequences, please visit our sequence

archive Web site (<http://www.fruitfly.org/sequence/>) or send email

to [bdg@fruitfly.berkeley.edu](mailto:bdg@fruitfly.berkeley.edu).

Pl library location: 67-43.

Location/Qualifiers

1. 77685

/organism="Drosophila melanogaster"

/strain="y2: cn bw sp"

/db.xref="taxon:7227"

/chromosome="2R"

/map="36D3-36D3"

/clone\_1lb="Pl library, partial Sau3A in pad10sacB1"

/clone="Pl DS06379 (D358)"

BASE COUNT 23941 a 16311 c 15292 g 22141 t

ORIGIN

Query Match 11.4% Score 41.8; DB 37; Length 77685;

Best Local Similarity 55.0%; Pred. No. 1.2;

Matches 82: Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 1 GCTCTTCTCTCCCTCCTGAAATTAATCTTTCACATTCGCAATTCGACAGATTACA 60

DB 24896 GATAATTCACGATGATGATCTTAAGTATGATTAATGATTAATGATTAAGAAATAAA 24837

OY 61 CATTTCACTGATGATATATGTGTGCAAAAAAAGTCTTCTTTAAATATAC 120

DB 24836 CACCTTGCATTAATTAATGATGTTTATTTACAAATAAACAATCTTTAATAAGTAC 24777

OY 121 TTGGTTGTGATCCATCTTCTTTTCC 149

DB 24776 TTACTTTATACCCAGTTTCTTGCCCC 24748

RESULT 6

AC006786 26236 bp DNA HTG 25-FEB-1999

LOCUS Caenorhabditis elegans clone Y4868, WORKING DRAFT SEQUENCE, 13

DEFINITION unordered pieces.

AC006786

NID 94309787

VERSION AC006786.2 GI:4309787

KEYWORDS HTG; HTGS; PHASE1.

SOURCE HMG; HTGS; PHASE1.

ORGANISM Caenorhabditis elegans.

Caenorhabditis elegans.

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilita; Rhabdilita;

Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita; Rhabdilita;

1 (bases 1 to 26236)

Waterston,R.H.

The sequence of Caenorhabditis elegans clone

Unpublished

2 (bases 1 to 26236)

Waterston,R.H.

Direct Submission

Submitted (23-FEB-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Mar 1, 1999 this sequence version replaced g1:4263152.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 13 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

COMMENT

Submitted (23-FEB-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

On Mar 1, 1999 this sequence version replaced g1:4263152.

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 13 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.



FEATURES	Source	Location/Qualifiers
1. 207370	/organism="Caenorhabditis elegans"	
	/db_xref="taxon:6239"	
	/clone="Y51F8"	
BASE COUNT	69298 a 36416 c 35867 g 65363 t 426 others	
ORIGIN		
Query Match	11.0%; Score 40.4; DB 34; Length 207370;	
Best Local Similarity	50.5%; Pred. No. 2.1;	
Matches	98: Conservative 0; Mismatches 96; Indels 0; Gaps 0.	
Oy	19 TCTGATTTAATTTCTTTCACATTCGCAATTTGCAAGGATTAACATTTCTCTGATGTA 78	
Db	148972 TCMAATTTAATTTTTCGAAAACATTTTGGCGGGAATTCAAATTTAATTTGTTGGAAA 148913	
Oy	79 ATGTGATTCGCAAAAAAAGAGTCCTTGTGTTAAATTAATTAAGTGTGTTGTAATTCATC 138	
Db	148912 ATATTTTGGCGGGAATTCAAACCTTTTAATTTTTCGAAAACATTTTGGCGGGAATTCAAAC 148853	
Oy	139 TTGCTTTTCCCATTTGSACTGACATGATTAATTAACCATCTGCACTGTAAGAAACATCT 198	
Db	148852 TTTAATTTTTCGAAAACATTTTACGGGGAATTCAAATTTAATTTTTCGAAAACATTT 148793	
Oy	199 GAAGACTGATCTA 212	
Db	148792 TGACGGGAATTCAA 148779	
RESULT	8	
AC006910/c		
LOCUS	AC006910 200509 bp DNA HTG 26-FEB-1999	
DEFINITION	Caenorhabditis elegans clone Y92H12, WORKING DRAFT SEQUENCE, 36	
ACCESSION	AC006910	
NID	94309895	
VERSION	AC006910.2 GI:4309895	
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	Caenorhabditis elegans.	
ORGANISM	Caenorhabditis elegans	
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;	
	Rhabditina; Rhabditoidae; Rhabditidae; Pelodierinae; Caenorhabditis	
	1 (bases 1 to 200509)	
	Waterston, R.H.	
	The sequence of Caenorhabditis elegans clone	
	Unpublished	
	2 (bases 1 to 200509)	
	Waterston, R.H.	
	Direct Submission	
	Submitted (24-FEB-1999) Genome Sequencing Center, Washington	
	University School of Medicine, 4444 Forest Park Parkway, St. Louis	
	MO 63108, USA	
COMMENT	On Mar 1, 1999 this sequence version replaced gi:4263438.	
	* NOTE: This is a 'working draft' sequence. It currently	
	* consists of 36 contigs. The true order of the pieces	
	* is not known and their order in this sequence record is	
	* arbitrary. Gaps between the contigs are represented as	
	* runs of N, but the exact sizes of the gaps are unknown.	
	* This record will be updated with the finished sequence	
	* as soon as it is available and the accession number will	
	* be preserved.	
	1 2013: contig of 2013 bp in length	
	* 2014 2028: gap of unknown length	
	* 2029 4358: contig of 2330 bp in length	
	* 4359 4373: gap of unknown length	
	* 4374 6822: contig of 2449 bp in length	
	* 6823 6837: gap of unknown length	
	* 6838 9471: contig of 2634 bp in length	
	* 9472 9486: gap of unknown length	
	* 9487 12251: contig of 2755 bp in length	
	* 12252 12266: gap of unknown length	
	* 12267 13524: contig of 3258 bp in length	

Query Match	Best Local Similarity	Score	DB	Length	200509
Matches	98	Conservative	0	Mismatches	96
				Indels	0
				Gaps	0

QY	19	TCGAAATTTAAATCTTTCACTCTGCATTTGCAAGATTTACACTGTCACATCGATGATTA	78
Db	194668	TCGAATTTTAAATTTTGTGAAAACATTTTGGCGGAAATTCAAATTTTAAATGTTTGGAAA	194609
QY	79	ATGTGTGTGCAAAAAAGTGTCTTGTGTTTAAATTAATCTGATGTTTGTGATCATC	138
Db	194608	AAATTTTGGCGGAAATTCAAACTTTAAATTTTGTGAAAACATTTTGGCGGAAATTCAAAC	194549
QY	139	TTGCTTTTTCCTCCCAATTGGAACTAGTCATTAACCATCTCTGAACCTGTAAGAAAAACATCT	198
Db	194548	TTTAATTTTGTGAAAACATTTTGGCGGAAATTCAAATTTTAAATTTTGTGAAAACATTTT	194489
QY	199	GAAGAGCTAGTCTA	212
Db	194488	TGACGGGAAATTCAA	194475

```

gene      complement(14133..36702)
          /gene="K12C11.4"
CDS       complement(3016(14133..14396,15675..16211,17274..17705,
          18554..19005,19953..20502,21661..21936,22047..22244,
          30994..31295,32682..33178,34086..34424,35131..35355,
          36467..36702))
          /gene="K12C11.4"
          /note="contains similarity to ankyrin repeats and protein
          kinases"
/codon_start=1
/evidence-not_experimental
/protein_id="AAB97579.1"
/db_xref="PID:g2804470"
/db_xref="GI:2804470"
/translation="MSTIALQQLLLLPYTLTRRSKSTKLKQNEGSPVVSAC
SPESKCEISCNITTFRIILNLAFLTIARQRIQGFVVRVRKTEKTAAFI
KRRVATSRGRGTROMIEREVRLQKIRGNSNVELHAYVETASDVIVLELVSGEL
FDHVAKECLEDEVAANAFIKOILAVRHLSHLIVHLDIRPENVMKRGDSQIKIID
FGLSREIEGAVKDVGTPEEVAPEVNVYEALESPATDMVAQVYVYILSGSPFLG
DNREDFTNTRVRYRHSRDSREKTSKAKDYFLRVFDVDORATVEEGLQHPWIRG
PENGADIRKASGITISHIOSPTROKRCVETLVVLLKASSSRIDGRDEDM
VASCTICABEGNLRALHKLSLKLPLNATRSKLSSEFSEPGATAMCAATYGHAE
VNFYFMKGGNICARDNGDPLHVACRFQHTVAVYVANEKIDVDSINKTGETALHC
AVESADTRVRLQLRPRLDLPNAREETPLHYAAARGVDCVQALLDANSPIDAVEQ
DEKTALEILENGNVDIASILITNGCDINHADHGTALHASKGLDANSPICHA
VWYDSVANRKTALHAHYGHVDIIRVLLARADYTLRGDGLTAELVAVAELEA
HSLKVRKSOEIREEYISOLYPRDTSLRKIKLGLHSOSGRRLVQTLHSSGICISF
LESVTRISDNHSPSSSKMDGTHSTNGSVSSNNSSFDLAAAGSRAPPHSQYT
RGIDVQTVINIGCGEVEFEGVEGEPHMCYDFVGNADCIHLILRYTDPPEVQKO
ILYWMNFKRGVYPEPIGHGFSRSRSYIIYGTHTSLSPFOMNOKEEYVSDIEA
MLNTVRLRETHVDHRLILDATNPSGKMTKMLAKCRITNLAKLEKTLID
TYVNHNLVLRKKANFPVITMPDFIOLVREINELTGDHRCROIYOOLIGFLVLR
NPLCDADYVLANEMEGTHIGOLLNABEELSRASPRNGSTSLATPPIPOSLM
TILEVLQCAPDARKGAHEPVITQTEADPSIDKRYSLKEKEDTVYGVRIIPMGK
ENSLHSTPRIOVALRSINDQAPADTOLHOMSECSKVSODREAVIRNMGVAVIR
ANGPSESATSMFYMEDLINLYEHAAEVPGISLRHFIKHEHPEHLPFPE
SMEMQRESLSVKTQDEBELFTDYVCGSRDVAHRLTGLDVGVAIDOMASRCLA
CLIDPHAGRDMSTLAVKLOLTDVODVDSGOSLRDOLLNEMAIHHPDQASVGN
LCRIIVELRCRDARDLRYRPLVYFAPLEDQDLFTNDGCVVSSCHSSEHPNINI"
repeat_region
/note="C. elegans telomere-like hexamer repeat"
25918..26212
BASE COUNT 13418 a 7528 c 7289 g 12244 t
ORIGIN

```

```

Query Match      11.08; Score 40.4; DB 36; Length 40479;
Best Local Similarity 50.5%; Pred. No. 2.8;
Matches 98; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

```

```

OY 19 TCTGAATTAACTTCACTTCATTTGCAAGATTACACTTCACATGATGAT 78
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24634 TCAAAATTTTATTTTGGAAAACATTTTGCGGGAAATTCAAATTTTATGTTTGA 24575
OY 79 ATTGTGTGCAAAAAAAGTCTCTTTTAAATTAATTCATGTTTGAATCCATC 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24574 ATATTTGGGGGAAATTCAACTTTATTTTGAAGAACATTTTGCGGGAATTC 24515
OY 139 TTGCTTTTCCCATTTGGAATCTAATACCATCTCGAAGTGTAGAAAAACATCT 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24514 TTAAATTTTGAAGAACATTTTAGCGGAATTCAAATTTTATTTTGAAGAACATTT 24455
OY 199 GAAGAGCTACTCTA 212
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 24454 TCAGCGGAATTCAA 24441

```

```

RESULT 10
DB7009/c 104386 bp DNA PRI 07-FEB-1999
LOCUS DB7009 104386 bp DNA PRI 07-FEB-1999
DEFINITION Human (lambda) DNA for Immunoglobulin light chain.
ACCESSION DB7009
NID 92114252
VERSION DB7009.1 GI:2114252

```

```

KEYWORDS 3'Oxyl.1; 5'Oxyl.1; V3-1P; V1-11; V4-1; V1-12P; Immunoglobulin
          light chain.
SOURCE Homo sapiens human pre-pro B cell line cell_line:FEb14-14 DNA,
          clone_lib:total human BAC library; Kelo BAC library clone:286A10.
ORGANISM Homo sapiens
          Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
          Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 104386)
AUTHORS Shimizu,N.
TITLE Direct Submission
JOURNAL Submitted (08-AUG-1996) to the DDBJ/EMBL/Genbank databases.
          Nobuyoshi Shimizu, Kelo University School of Medicine, Department
          of Molecular Biology; 35 Shinanomachi, Shinjuku-ku, Tokyo 160,
          Japan (E-mail:shimizu@med.kelo.ac.jp, Tel:03-3351-2370,
          Fax:03-3351-2370)
2 (sites)
REFERENCE 2 (sites)
AUTHORS Kawasaki,K., Minoshima,S., Nakato,E., Shibuya,K., Shintani,A.,
          Schmeits,J.L., Wang,J. and Shimizu,N.
TITLE One-megabase sequence analysis of the human Immunoglobulin lambda
          gene locus
JOURNAL Genome Res. 7 (3), 250-261 (1997)
MEDLINE 97228902
FEATURES
          Location/Qualifiers
            1..104386
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /cell_line="FEb14-14"
              /cell_type="human pre-pro B cell line"
              /chromosome="22"
              /clone="286A10"
              /clone_lib="total human BAC library; Kelo BAC library"
              /map="22q11.2"
            1..104386
              /note="region III"
              /gene="V1-12P"
              /gene="V1-12P"
              /gene="V1-12P"
              /gene="V1-12P"
              /pseudo
              /codon_start=1
              /codon_start=1
              /gene="V4-1"
              /gene="V4-1"
              /gene="V4-1"
              /gene="V4-1"
              /codon_start=1
              /codon_start=1
              /protein_id="BAA20003.1"
              /db_xref="PID:g1020804"
              /db_xref="PID:g2114253"
              /db_xref="GI:2114253"
              /translation="QPVLTQPPSSASRPSGARLCTLPSPDINGSYNITYYQQRGS
              PPRILYISDSKDGSGVSPRFSKDSANTGILLISGLQSEDEADYICAMWDSLN"
              AS"
              13275..13566
              /gene="V1-11"
              /gene="V1-11"
              /gene="V1-11"
              /gene="V1-11"
              /codon_start=1
              /codon_start=1
              /protein_id="BAA20004.1"
              /db_xref="PID:g1020805"
              /db_xref="PID:g2114254"
              /db_xref="GI:2114254"
              /translation="QSVLTQPPSVSEAPRQRYTISGSGSSNIGNAAVWYQOLPEKA
              PKLITYDDLPSGVSDFSGSGTASLASISGLQSEDEADYICAMWDSLN"
              31459..31753
              /gene="V3-1P"
              /gene="V3-1P"
              /gene="V3-1P"
              /gene="V3-1P"
              /pseudo
              /codon_start=1
              /codon_start=1
              /complement(68859..70486)
              /gene="5'Oxyl.1"
              /complement(68859..70486)
              /gene="5'Oxyl.1"

```

```

/note="putative"
/codon_start-3
/protein_id="BAA20005.1"
/db_xref="PID:d1020806"
/db_xref="PID:q2114255"
/db_xref="GI:2114255"
/translation="OSCEBEKPEPOKNIQETKQVDDDEADLIFVGEVHNEADLIF
VGVSTNSRPNVSNILNRYTPGSMSSRRKDYDLKDKTARKLQKSHETVTSKAVTVLPA
SOLESRTDPSPIIEPLKDYRNSPOVYVNSSELPLITFDLSLHHPVSTALSV
GGINSPRQKSLSTFEVNSINPKRAKLADGIIENSSASPSPDQPHMTNQQSPSN
NVHTLSHYQNSAPRPPAARPKRNINIKNTLIDDEBNELAKTIDILSTSONKTPDKP
ENPIVLSDFTIGHKHGBQPEOKTHFTFKCLSCVAKLVKNVFNHVAHHEFEQRN
DSMENVHTCOHCHQRPPTPEFLOCHIENVTQAEPSYVCKICELSEPLKIFKTAATPYMCHYR
HKKPEMPYVCOVCHYRSSFVADVEHETFCHEKNTKILCPCLKIFKTAATPYMCHYR
HGMSKSAOCSKRLQFLTEFKRMEHKTQCHQMFKPKQLEGLPPEVTIVTSLEPL
OPGSYDVASITVSTSDSEPSLRKSKISKSH"
complement(95093..6717)
/gene="3.0Y11.1"
/complement(95093..>96717)
/gene="3.0Y11.1"
/note="putative"
/codon_start-3
/protein_id="BAA20006.1"
/db_xref="PID:d1020807"
/db_xref="PID:q2114256"
/db_xref="GI:2114256"
/translation="DIFLCRKVESPKNLRSEKOREDEDEPDLIYGVGEVHNRDAEV
LEFGMINSKPYVSNILNRYTPGSMSSRRKDYDLKDKTARKLQKSHETVTSKAVTVLPA
LSEGRSTDSPTVMKSSSEPGYKMSPOVYVNSSELPLITFDLSLHHPVSTALSV
PDSKRLSTSDINSRDSKRYKLDGIPVPSLAVPSDMSSTISTVTPSQGICNSSNHV
ONGVTFPMPDANGKAHNLTPDERANESGLANTDLSLASONKTPDKRENITVLNS
FYGOHKHGBQPEOKTHFTFKCLSCVAKLVKNVFNHVAHHEFEQRNDSMENVHTCO
HCHQRPPTPEFLOCHIDSVHJAMGPSAVCKICELSEPLKIFKTAATPYMCHYR
VCOVGHSSVADVADVEHETFCHEKNTKILCPCLKIFKTAATPYMCHYRHSRRAVIO
CSKRLQFLTEFKRMEHKTQCHQMFKPKQLEGLPPEVTIVTSLEPL
STVDQSSVVKTKKTKAMTRNSRLPCSKDSS"
BASE COUNT      31195 a 22463 c 22062 g 28666 t
ORIGIN
Query Match      10.9% Score 39.8; DB 9; Length 104386;
Best Local Similarity 57.7% Pred. No. 3.2;
Matches 71; Conservative 0; Mismatches 52; Indels 0; Gaps 0;
QY      5 TTTCTCTCCCTCTGATTAATTCCTTCACTGCAATTGCAAGATTACACATT 64
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 51580 TTTCTTCCCTTTCCTGCAAGTCTCTTTATCTGACATTTAATACCCACTGTAT 51521
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY      65 TCACGTGATGATATGTTGCTGCAAAAAAAGTCTTTGTTAAATTAATTACTGG 124
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 51520 TTTCTTGAATGAGATGTTGAGAAATATTAATTTAATCTTAATTAATTAGTTT 51461
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
QY      125 TTT 127
      ||
Db 51460 ATT 51458

```

```

RESULT 11
CEK08E3 LOCUS      CEK08E3      39565 bp      DNA      INV      23-NOV-1998
DEFINITION Caenorhabditis elegans cosmid K08E3, complete sequence.
ACCESSION  Z81568
NID          93643297
VERSION      281568.1 GI:3643297
KEYWORDS     HTG.
SOURCE       Caenorhabditis elegans.
ORGANISM     Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditiia; Rhabditiida;
Rhabditiina; Rhabditiidae; Rhabditiidae; Peloderinae; Caenorhabditis.
REFERENCE    1 (bases 1 to 39565)
AUTHORS      McMurtry,A
TITLE        Direct Submission
JOURNAL      Submitted (06-NOV-1996) Louls, MO 63110, USA. E-mail:

```

```

REFERENCE
AUTHORS      jes@sanger.ac.uk or rwenematode.wustl.edu
2 (bases 1 to 39565)
WILSON,R., Ainscough,R., Anderson,K., Baynes,C., Berts,M.,
Bonfield,J., Burton,J., Connell,M., Cosey,T., Cooper,J.,
Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A.,
Fullon,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M.,
Johnston,L., Jones,M., Kersey,J., Kisters,J., Laister,N.,
Latreille,P., Lightning,J., Lloyd,C., McMurtry,A., Mortimore,B.,
O'Callaghan,M., Parsons,J., Percy,C., Riffen,L., Roopra,A.,
Saunders,D., Showkneen,R., Smalton,N., Smith,A., Sonhammer,E.,
Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M.,
Vaughan,K., Waterston,R., Watson,A., Weinstock,L.,
Wilkinson-Spoat,J. and Woldman,P.
2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans
JOURNAL
MEDLINE
COMMENT
On Sep 23, 1998 this sequence version replaced gi:3294143.
Coding sequences below are predicted from computer analysis, using
predictions from Genefinder (P. Green, U. Washington), and other
available information.
For a graphical representation of this sequence and its analysis
see:
http://webc.sanger.ac.uk/cgi-
bin/display?db=wormaceclass=sequence&object=K08E3
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
This sequence is the entire insert of clone K08E3. The true right
end of clone W06F12 is at 467 in this sequence. The start of this
sequence (1..105) overlaps with the end of sequence Z83244.
The end of this sequence (39514..39565) overlaps with the start of
sequence AL032671.
FEATURES
source
gene
CDS
location/Qualifiers
1..39565
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="III"
/clone="K08E3"
3510..6358
/gene="K08E3.1"
join(3510..3749,3844..4023,4219..4319,4374..4816,
4998..5101,5288..5683,5730..6167,6215..6358)
/gene="K08E3.1"
/note="similar to Common central domain of tyrosinase;
cDNA EST EMBL:Z14371 comes from this gene; cDNA EST
CEMSG81P comes from this gene; cDNA EST EMBL:M69185 comes
from this gene; cDNA EST EMBL:C13572 comes from this gene;
cDNA EST YK240a11.3 comes from this gene; cDNA EST
YK240a11.5 comes from this gene"
/codon_start=1
/protein_id="CAB04594.1"
/db_xref="PID:e1348001"
/db_xref="PID:g3878440"
/db_xref="GI:3878440"
/translation="MRWTPILTTLLCQVLIISVAGVSTTKKAGVGTAVRKTIVK
VAVKNKIQRMKNRMRPPDCSDARHAKOTCLMRDLARRLARLARSQVOTSP
NOYVNWLOPIVPPANAGQAAYHPHYDCITLCLCPFNRRGVNQCYSLVNMSM
KRETRMTEDERRRHHNLNLRNGEDIRLSROHLEVGSGASHSPGLPWREYL
KRVEIALRMADPTVEFIPWDSVMDSYLPDPDSIMFSDLPVGGTDYGNVTPFAWM
RTIEGRSTILRLNLAEGQLFENQVNTIVAQNTIENTILAYTAPQPCPYPNVYCAIEY
TSNINHLIMIGADMKPPSTSAHEPIEFMHSHVDVITLWELMROLOQPRMLREOAYSADHY
TCAMWQHSTVAMPAPPIVAVRGLINSYSDWRYAPRATCSORBNQSPYFCOT
RGYHCYSKILNLCNCRGFEDNDACYSRCMGCVANFPAARKGARTHANSTLIDL
ELAKEVSYNGKTVRVERPLVKRISIKMTAPLEFVDCYNMPCDSSWAKNGCGOREFYS
QYCAQSNCTPSYNTDADMDRVNCAVMKADMGKSGKSDPSEKSCGCGCSLT
RESOCFNNGKIIVPSONLRESISDSIMSEVTNRIRIKVKVAKEEVESTKVRHN"

```



gene  
complement(6755. .7969)  
/gene="K08E3.2"  
complement(join(6755. .6953,7015. .7141,7190. .7388,  
7618. .7680,7727. .7969))  
/gene="K08E3.2"  
/note="predicted using GeneFinder"  
/codon\_start=1  
/protein\_id="CAB04590.1"  
/db\_xref="PID:e1347997"  
/db\_xref="PID:g3878436"  
/db\_xref="GI:3878436"  
/translation="MEYEVLYVALISLCSLIIGQCTSKKAPVAPIMGEKHAESLTA  
TPVSPAPVSSSLALDEQDEKKKEGEEKEEKDEKDEKDEKDEKDEKDEK  
EDHEKKKAPVSEENNEKKDENKKRDKLKKKAKWFKCDSVAVITINPFGK  
DDKKAKMSVDAPELIFERTITSQRLKKLKKKAKWFKCDSVAVITINPFGK  
IDPMSVTAETVITIHRSKEQEDKLVTPAKMKGELEMAQAFQRIKKTGSALTVALVSS  
".

gene  
14167. .18631  
/gene="K08E3.3a"  
14167. .18631  
/gene="K08E3.3b"  
join(14167. .14295,14596. .14763,14810. .14911,15156. .15314,  
15357. .15472,15522. .15645,15694. .15854,15902. .16007,  
16055. .16259,16305. .16386,16433. .16509,16656. .16905,  
17201. .17352,17487. .17593,17642. .17692,18137. .18266,  
18350. .18524,18568. .18631)  
/gene="K08E3.3b"  
/note="similar to EF hand, Src homology domain 3; CDNA EST  
YK273e3.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB04595.1"  
/db\_xref="PID:e1348002"  
/db\_xref="PID:g3878441"  
/db\_xref="GI:3878441"  
/translation="MEEKVASTCFHWTLTKDYIVFYVSSNNLNFASRMPRRRGORLL  
PALRAMINPEALAEVSKARTSGARTVEEPIPIKKYEAAGCRSTLKEQOTLL  
SHFDREGNOIIMLEKSLMLONGCEKMTNOERALKAPATGDLTIRLNINFNKTTLDH  
SAIGLRLTQVDMIPVSRFTVODPSNALSTYQKIDFEFKLQSEKKAIEE  
EYSLRLSTAKRYARKSEEDDEILKSVSTSSFNFLQDLQIARHQTSEHJNGV  
VYVASKTCOMRSKRKNALNDKTLNDKLEQIDNMGKSGKCYLRFVDAESYQFV  
KADNKLSEIRLEKARALANARNKECAEQDYSALVRYKTAEOKRKAFVDAEPIYFA  
RKAYDECIADMRQVLQKTVSFDSIADSPCECKIMQREPVKIKDAEGDQIVKSV  
EKTIEDPAFETIEDQDFKNCDSRTNDSADSGGKULLSSPKNKILINFTGLKEKE  
ADEPEKSNNDQMLTYDKSRPAHVRSLCRSLRDMERQLEQALIGRGITRLQDAVY  
TTPHGNPACTEPLISYAKKIEKLMQIHNFKEYAMLEMSVEGQERSGGRDTP  
TRNSGSSSTNOSSKTTIEDVSGAGNSSADSSKNILROLFTTPKRLISPTSK  
SSTPIPLRRRAETSSPKILRSGCAIRKSLSDPSVETRAVTALEFPKASSAET  
MSIEGEELIVLEHHDGDMWTRKNCRKRNHESGVPFYSIDFPO"  
join(14167. .14295,14596. .14763,14810. .14911,15156. .15314,  
15357. .15472,15522. .15645,15694. .15854,15902. .16007,  
16055. .16259,16305. .16386,16433. .16509,16656. .16905,  
17201. .17352,17487. .17593,17642. .17692,18137. .18266,  
18350. .18524,18574. .18631)  
/gene="K08E3.3b"  
/note="predicted using GeneFinder; similar to EF hand, Src  
homology domain 3; CDNA EST EMBL:D64948 comes from this  
gene; CDNA EST EMBL:D66852 comes from this gene; CDNA EST  
EMBL:D76408 comes from this gene; CDNA EST EMBL:D70789  
comes from this gene; CDNA EST EMBL:D68217 comes from this  
gene; CDNA EST EMBL:C13458 comes from this gene; CDNA EST  
YK273e3.3 comes from this gene"  
/codon\_start=1  
/protein\_id="CAB04591.1"  
/db\_xref="PID:e1347998"  
/db\_xref="PID:g3878437"  
/db\_xref="GI:3878437"  
/translation="MEKVASTCFHWTLTKDYIVFYVSSNNLNFASRMPRRRGORLL  
PALRAMINPEALAEVSKARTSGARTVEEPIPIKKYEAAGCRSTLKEQOTLL  
SHFDREGNOIIMLEKSLMLONGCEKMTNOERALKAPATGDLTIRLNINFNKTTLDH  
SAIGLRLTQVDMIPVSRFTVODPSNALSTYQKIDFEFKLQSEKKAIEE  
EYSLRLSTAKRYARKSEEDDEILKSVSTSSFNFLQDLQIARHQTSEHJNGV  
VYVASKTCOMRSKRKNALNDKTLNDKLEQIDNMGKSGKCYLRFVDAESYQFV  
KADNKLSEIRLEKARALANARNKECAEQDYSALVRYKTAEOKRKAFVDAEPIYFA  
RKAYDECIADMRQVLQKTVSFDSIADSPCECKIMQREPVKIKDAEGDQIVKSV  
EKTIEDPAFETIEDQDFKNCDSRTNDSADSGGKULLSSPKNKILINFTGLKEKE  
ADEPEKSNNDQMLTYDKSRPAHVRSLCRSLRDMERQLEQALIGRGITRLQDAVY  
TTPHGNPACTEPLISYAKKIEKLMQIHNFKEYAMLEMSVEGQERSGGRDTP  
TRNSGSSSTNOSSKTTIEDVSGAGNSSADSSKNILROLFTTPKRLISPTSK  
SSTPIPLRRRAETSSPKILRSGCAIRKSLSDPSVETRAVTALEFPKASSAET  
MSIEGEELIVLEHHDGDMWTRKNCRKRNHESGVPFYSIDFPO"

CDs

	gene						
CDS		<p>/gene="K08E3.4"</p> <p>/complement(join(18904..19472,19520..20118,20164..20373, 20433..20734,21052..21164,21214..21352))</p> <p>/gene="K08E3.4"</p> <p>/note="Predicted using GeneFINDER: similar to Cofilin/tropomyosin-type actin-binding proteins, Src homology domain 3; CDNA EST Yk442b7.5 comes from this gene; CDNA EST EMBL:D14844 comes from this gene; CDNA EST EMBL:M88991 comes from this gene; CDNA EST EMBL:D14502 comes from this gene; CDNA EST EMBL:D71424 comes from this gene; CDNA EST EMBL:D74120 comes from this gene; CDNA EST EMBL:D71594 comes from this gene; CDNA EST EMBL:D74252 comes from this gene; CDNA EST EMBL:D72634 comes from this gene; CDNA EST EMBL:D68488 comes from this gene; CDNA EST Yk207d8.3 comes from this gene; CDNA EST Yk207d8.5 comes from this gene; CDNA EST Yk274b9.3 comes from this gene; CDNA EST Yk274b9.5 comes from this gene; CDNA EST Yk328d11.3 comes from this gene; CDNA EST Yk328d11.5 comes from this gene; CDNA EST Yk335b5.3 comes from this gene; CDNA EST Yk335b5.5 comes from this gene; CDNA EST Yk369d4.3 comes from this gene; CDNA EST Yk369d4.5 comes from this gene; CDNA EST Yk376a8.3 comes from this gene; CDNA EST Yk376a8.5 comes from this gene; CDNA EST Yk399a9.3 comes from this gene; CDNA EST Yk399a9.5 comes from this gene; CDNA EST Yk419c7.3 comes from this gene; CDNA EST Yk442b7.3 comes from this gene; CDNA EST Yk462c7.3 comes from this gene; CDNA EST Yk462c7.5 comes from this gene; CDNA EST Yk465d12.3 comes from this gene; CDNA EST Yk465d12.5 comes from this gene"</p> <p>/codon_start=1</p> <p>/protein_id="CAB04592.1"</p> <p>/db_xref="PID:e1347999"</p>					
Query Match	10.8%	Score 39.6:	DB 36:	Length 39565:			
Best Local Similarity	51.1%:	Pred. No. 4.2:					
Matches 93:	Conservative 0:	Mismatches 89:	Indels 0:	Gaps 0:			
Oy	23	AATTAAATCTTTCAACTGCAATTTGGCAAGATTACACATTCACGTGATATATTG	82				
Db	8505	ATTCAAATTTTCTGGAATAATATTTTGCGGGGAATTCAAATTTTAATTTTTAAAAAATAT	8564				
Oy	83	TGTTGCACAAAAAAAAGTGCTCTTGTGTTAAAAATTACTTGTTGTGTAATCCATCTTGC	142				
Db	8565	TTTGCGGGGAATTAATAATTTTAATTTTGAATAATATTTTGCGGGGAATTAATAATTTT	8624				
Oy	143	TTTTTCCCCCATGGACAGTACATTAACCACATCTGTGACGTGAGAAAAACAATCTGAAG	202				
Db	8625	AATTTTGGAAAAATATTTTGCGGGGAATTCAAATTTTAATTTTGGAAAAATATTTTGCGG	8684				
Oy	203	AG 204					
Db	8685	CG 8686					
RESULT 12							
LOCUS AC006710	AC006710	145920 bp	DNA	HTG	23-FEB-1999		
DEFINITION Caenorhabditis elegans c1one Y119C1AA, WORKING DRAFT SEQUENCE, 1							
unordered pieces.							
ACCESSION AC006710							
NID 94263228							
VERSION AC006710.1 GI:4263228							
KEYWORDS HTG; HTGS_PHASE1.							
SOURCE Caenorhabditis elegans.							
ORGANISM Caenorhabditis elegans							

REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis
AUTHORS	1 (bases 1 to 145920)
TITLE	Waterston, R.H.
JOURNAL	The sequence of <i>Caenorhabditis elegans</i> clone
REFERENCE	Unpublished
AUTHORS	2 (bases 1 to 145920)
TITLE	Waterston, R.H.
JOURNAL	Direct Submission
COMMENT	Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
FEATURES	* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. * 1 145920: contig of 145920 bp in length. * Location/Qualifiers 1..145920 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /clone="Y119C1AA"
SOURCE	
BASE COUNT	46442 a 26006 c 25762 g 47710 t
ORIGIN	
Query Match	10.6%; Score 38.8; DB 34; Length 145920;
Best Local Similarity	48.9%; Pred.No. 5;
Matches 134; Conservative	0; Mismatches 137; Indels 3; Gaps 1;
Oy 21	TGAATTAATTCTTTCACACTTGCAAGATTACACATTCACGTGATGAT 80
Dn 129578	TCAATTTCTCTTAAACGGGTAATTTTCGCAAAAAAATCAATTAATAAAC 129637
Oy 81	TGTGTGCACAAAAGAGTCTTGTGTTAAATTAATCTGGTGTGAATCCATCTT 140
Dn 129638	CAAATTTCTTAAAGTCGAATTTTTGTGTTGTAATAAGTAGTACTTTATAGAGAAATG 129697
Oy 141	GCTTTTCCCCCTTGGAACTACTCATTAACCATCTCTCACTGGTAGAANAACATCTCA 200
Dn 129698	TGTAATTTTCCCCTCGACATCAAAAAATTTGGAAGAT--TGAATTTTAAATCAATTTCT 129754
Oy 201	AGAGCTAGCTCATCGACATCTGACAGGTGAATGGATGGTCTCGAACCATTTCACCCA 260
Dn 129755	AAATGGAACACTTTCGTTTTCATATTTTATTTTAAAAAATGGAANAACGATACCCTTG 129814
Oy 261	GACAGCCCTGTTCTATCCTGTTTAAATTAATTAAGT 294
Dn 129815	ATTGAATATTTTGTGTTTGAATTAATAAACCTAGT 129848
RESULT 13	
LOCUS	AC006712 299919 bp DNA HTG 23-FEB-1999
DEFINITION	<i>Caenorhabditis elegans</i> clone Y119c1b, WORKING DRAFT SEQUENCE, 2
ACCESSION	unordered pieces.
NID	AC006712
VERSION	94263226
KEYWORDS	AC006712.1 GI:4263226
SOURCE	HTG; RTGS_PHASE1.
ORGANISM	<i>Caenorhabditis elegans</i> .
REFERENCE	<i>Caenorhabditis elegans</i> .
AUTHORS	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdilia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis
TITLE	1 (bases 1 to 299919)
JOURNAL	Waterston, R.H.
REFERENCE	The sequence of <i>Caenorhabditis elegans</i> clone
UNPUBLISHED	Unpublished
REFERENCE	2 (bases 1 to 299919)

AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT	* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved * 1 128484: contig of 128484 bp in length * 128485 128499: gap of unknown length * 128500 299919: contig of 171420 bp in length. Location/Qualifiers 1..299919 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /clone="Y119C1b"
FEATURES	
source	
BASE COUNT	96329 a 53231 c 54548 g 95796 t 15 others
ORIGIN	
Query Match	10.6%; Score 38.8; DB 34; Length 299919; Best Local Similarity 48.9%; Pred.No.4.5;
Matches 134; Conservative 0; Mismatches 137; Indels 3; Gaps 1;	
OY 21 TGAATTATTTCTTCAACTTCGACATTTTGCAAGGTTCACACTTCACGCGATGTAT 80	
Ds 258077 TCAATTTTCTTTAAACGGGTAAATTTTCGCAAAAAAACAATTAATAAAAAAAC	258136
OY 81 TGCTGTGCAAAAAAAAAAAGTGTCTTTGTTTAAATTACTTGTTGTGAATCCATCTT 140	
Ds 258137 CAATTTCTAAAGTCGATTTTGTGTGTAANAATAAGTACTTTTATAAGAGATTG 258196	
OY 141 GCTTTTCCCCTTGGAACCTAGTCATTTAACCATCTCTGCAGACTGGTAGAAAACATCTGA 200	
Ds 258197 TGTATTTTCCCATCGACATTAATAAATTGGAAGAT--TGAATTTTTTAATCAATTTTCT 258253	
OY 201 AGAGCTAGTCATTCAGCATTCGACAGGGAATGATGGTTCTCAGAACCATTTTACCCA 260	
Ds 258254 AATCGAACAATTCCTGTTTTCATATTTTATTTTAAAAAATCGAAAAAGCATACCCTTG 258313	
OY 261 GACAGCCGTGTTCTCTANCCGTTTAAATTAATAGT 294	
Ds 258314 ATTAGAATATTTTGTGTTTGGATTAATAAACCTAGT 258347	
RESULT 14	
CEY11B2_0	
MPCOMMENT	
Sequence split into 7 fragments LOCUS CEY11B2 Accession Z98857	
Fragment Name Begin End	
CEY11B2_0 1 110000	
CEY11B2_1 100001 210000	
CEY11B2_2 200001 310000	
CEY11B2_3 300001 410000	
CEY11B2_4 400001 510000	
CEY11B2_5 500001 610000	
CEY11B2_6 600001 613487	
LOCUS CEY11B2 613487 bp DNA HTG 29-MAY-1999	
DEFINITION Caenorhabditis elegans chromosome unknown clone Y11B2, WORKING	
DRAPT SEQUENCE, In unordered pieces.	
Z98857	
NTD 94938505	
VERSION 298857.32 GI:4938505	
KEYWORDS HTG, RTGS, PHASE1.	
SOURCE Caenorhabditis elegans.	
ORGANISM Caenorhabditis elegans	
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida; Rhabditoidea; Rhabdificae; Peloderinae; Caenorhabditis	



---

...



## RESULT 2

V61203  
 ID V61203 standard; cDNA: 366 BP.  
 AC V61203:  
 DT 06-JAN-1999 (first entry)  
 DE cDNA sequence of prostate tumour clone.  
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
 OS Homo sapiens.  
 PN WO9837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998; U03492.  
 PR 09-FEB-1998; US-020956.  
 PR 25-FEB-1997; US-806099.  
 PR 01-AUG-1997; US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI: 98-609886/51.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 used in a vaccine for the treatment of prostate cancer  
 PS Claim 3; Page 85; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 can be used as a vaccine for the treatment of prostate cancer. The DNA  
 was identified by analysis of a subtracted cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SQ Sequence 366 BP: 102 A; 78 C; 58 G; 128 T;

Query Match 100.0%; Score 366; DB 1; Length 366;  
 Best Local Similarity 100.0%; Pred. No. 9.2e-90;  
 Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 GCCTTCTCTCCCTCCCTGATTAATTTCTTCACTGCAATTTGCAAGGATTACA 60
    |||
DB 1 GCTCTTCTCTCCCTCCCTGATTAATTTCTTCACTGCAATTTGCAAGGATTACA 60
OY 61 CATTCTACTGTGATATATTTGTTGCAAAAAAAGTGTCTTTGTTAAATTAC 120
    |||
DB 61 CATTCTACTGTGATATATTTGTTGCAAAAAAAGTGTCTTTGTTAAATTAC 120
OY 121 TTGCTTTGGAATCCATCTTGTCTTTTCCCATGGAAGTCAATCCATCTCTGA 180
    |||
DB 121 TTGCTTTGGAATCCATCTTGTCTTTTCCCATGGAAGTCAATCCATCTCTGA 180
OY 181 ACTGGTAGAAAAACATCTGAAGAGTACTGATCTATCAGCATCTGACAGTAAATGGATGT 240
    |||
DB 181 ACTGGTAGAAAAACATCTGAAGAGTACTGATCTATCAGCATCTGACAGTAAATGGATGT 240
OY 241 TCTCAGAACCATTTTACCCAGACAGCCTGTTCTATCTGTTTAATTAATTAATTTGGGT 300
    |||
DB 241 TCTCAGAACCATTTTACCCAGACAGCCTGTTCTATCTGTTTAATTAATTAATTTGGGT 300
OY 301 TCTCTACATGATACAAACCTGCTCCATCTGTGACATTAAGTCTGTGACTGAGT 360
    |||
DB 301 TCTCTACATGATACAAACCTGCTCCATCTGTGACATTAAGTCTGTGACTGAGT 360
OY 361 TTAGTC 366
    |||
DB 361 TTAGTC 366

```

## RESULT 3

V58614/c  
 ID V58614 standard; cDNA: 335 BP.  
 AC V58614:  
 DT 08-DEC-1998 (first entry)  
 DE Prostate tumour specific gene clone.  
 KW Prostate tumour specific gene; human; prostate cancer; detection;  
 therapy; ss.  
 OS Homo sapiens.  
 PN WO9837418-A2.  
 PD 27-AUG-1998.

PF 25-FEB-1998; U03690.  
 PR 09-FEB-1998; US-904809.  
 PR 25-FEB-1997; US-806596.  
 PR 01-AUG-1997; US-904809.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI: 98-480805/41.  
 PT Novel human prostate specific tumour protein and fragments - useful  
 for detecting and treating prostate cancers  
 PS Claim 1; Page 100; 141pp; English.  
 CC This sequence represents a human prostate tumour specific gene, and can  
 be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprises contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC encoded by this sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 SQ Sequence 335 BP: 117 A; 55 C; 66 G; 94 T;

Query Match 72.4%; Score 265; DB 1; Length 335;  
 Best Local Similarity 98.9%; Pred. No. 1.2e-62;  
 Matches 265; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

OY 99 AAGTCTCTTTGTTAAATTTACTGTTGTAATCCATCTTGTCTTTCCCATGGAA 158
    |||
DB 335 AAGTCTCTTTGTTAAATTTACTGTTGTAATCCATCTTGTCTTTCCCATGGAA 276
OY 159 CTAGTCATTAACCCATCTCTGACTGTAGAAAAACATCTGAGAGTACTATACCA 218
    |||
DB 275 NTAGTCATTAACCCATCTCTGACTGTAGAAAAACATCTGAGAGTACTATACCA 216
OY 219 TCTGACAGTGAATGATGTTCTCAGAACCAATTTACCCAGACAGCCTTTCTATCC 278
    |||
DB 215 TCTGACAGTGAATGATGTTCTCAGAACCAATTTACCCAGACAGCCTTTCTATCC 156
OY 279 TGTAAATTAATTAATTTGTTGGTCTCTCATGCAATCAAAACCTGCTCCATCTGTAC 338
    |||
DB 155 TGTAAATTAATTAATTTGTTGGTCTCTCATGCAATCAAAACCTGCTCCATCTGTAC 96
OY 339 ATAAAGTCTGTGAAGTTAGTC 366
    |||
DB 95 ATAAAGTCTGTGAAGTTAGTC 68

```

## RESULT 4

V61229/c  
 ID V61229 standard; cDNA: 335 BP.  
 AC V61229:  
 DT 06-JAN-1999 (first entry)  
 DE cDNA sequence of prostate tumour clone.  
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
 OS Homo sapiens.  
 PN WO9837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998; U03492.  
 PR 09-FEB-1998; US-020956.  
 PR 25-FEB-1997; US-806099.  
 PR 01-AUG-1997; US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI: 98-609886/51.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 used in a vaccine for the treatment of prostate cancer  
 PS Claim 3; Page 94; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
 CC was identified by analysis of a subtracted cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.

Sequence 335 BP: 117 A: 55 C: 66 G: 94 T:

Query Match 72.4%; Score 265; DB 1: Length 335;  
 Best Local Similarity 98.9%; Pred. No. 1.2e-62;  
 Matches 265; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

99 AAGTCTCTTTGTTAAATTAATCTGTTGTTGTAATCCATCTGTTTCCCATGGA 158  
 |||||  
 DB 335 AAGTCTCTTTGTTAAATTAATCTGTTGTTGTAATCCATCTGTTTCCCATGGA 276

159 CTAGTATTAAACCACTCTGAACTGGTAGAAAAACATCTGAAGAGCTAGTATACGA 218  
 |||||  
 DB 275 NTAGTATTAAACCACTCTGAACTGGTAGAAAAACATCTGAAGAGCTAGTATACGA 216

219 TCTGACAGGTGAATGGATGTTCTCGAAGCAATTTACCCAGACAGCTGTTTATCC 278  
 |||||  
 DB 215 TCTGACAGGTGAATGGATGTTCTCGAAGCAATTTACCCAGACAGCTGTTTATCC 156

279 TGTATAATAATTAGTTGGTCTCTACATGCATACAAACCTGCTCCATCTGTAC 338  
 |||||  
 DB 155 TGTATAATAATTAGTTGGTCTCTACATGCATACAAACCTGCTCCATCTGTAC 96

339 ATAAAGTCTGTGACTTGAAGTTAGTC 366  
 |||||  
 DB 95 ATAAAGTCTGTGACTTGAAGTTAGTC 68

RESULT 5  
 ID V74409/C  
 AC V74409:  
 DT 16-MAR-1999 (first entry)  
 DE Staphylococcus aureus contig SEQ ID #98.  
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;  
 KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.  
 OS Staphylococcus aureus  
 FH Key  
 FT misc-feature  
 541..600  
 /tag- a  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

2341..2400  
 /tag- b  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

4141..4200  
 /tag- c  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

5941..6000  
 /tag- d  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

7741..7800  
 /tag- e  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

9341..9600  
 /tag- f  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

misc-feature  
 9341..9600  
 /tag- f  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

EP-786519-A2.  
 30-JUL-1997.  
 PF 07-JAN-1997; 100117.  
 PR 05-JAN-1996; US-009861.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Barsh SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA, Rosen CA,  
 PI WPI: 97-374922/35.  
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus  
 PT stored on computer readable medium and used in the production of  
 PT anti-S.aureus vaccines

PS Claim 1: Page 584-590; 327pp; English.  
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences  
 CC of the invention. The DNA sequences are recorded on a computer readable  
 CC medium, preferably selected from a floppy or hard disk, random access  
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
 CC the S.aureus DNA sequences allows putative functions to be assigned so  
 CC that protein-encoding or regulatory regions of commercial, therapeutic or  
 CC industrial importance can be obtained. Specifically, sequences which are  
 CC likely to encode antigens have been identified and these polypeptides can  
 CC be used in a vaccine composition against S.aureus infection. The  
 CC polypeptides can also be used in a kit for the immunodetection of  
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,  
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
 CC for recombinant production of the polypeptides. The new DNA sequences  
 CC (and their fragments) are useful as primers or probes for isolating  
 CC homologues of any of the S.aureus DNA sequences contained on the  
 CC computer readable medium.  
 CC Sequence 10322 BP: 3655 A: 1302 C: 1989 G: 3014 T:

Query Match 10.4%; Score 38; DB 1: Length 10322;  
 Best Local Similarity 55.2%; Pred. No. 0.28;  
 Matches 74; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

14 CCTCTCTGAAATTAATTTTCACATTCGCAAGATTACACATTCACCTGTA 73  
 |||||  
 DB 9394 CGTATCTTAATATGATGATTAACATAAGCAATTAATTAATTAATTTG 9335

74 TGTATTTGTGTGCAAAAAAAGTGTCTTTTAAATTAATCTGTTGTAAT 133  
 |||||  
 DB 9334 TTTTGTGTTTAAATTAATTAAGTAATTTCTCATCTAATTAATTAATTAAT 9275

134 CCATCTGCTTTT 147  
 |||||  
 DB 9274 GTATTGTGTTAT 9261

RESULT 6  
 ID 021188  
 AC 021188:  
 DT 16-MAY-1992 (first entry)  
 DE Sequence encoding a 302 amino acid endochitinase which lacks a 3'  
 DE region, from plasmid pCH3.5.  
 KW Pest-resistant plant; transformed plant; fungi; insect; bacteria;  
 OS Tomato.  
 FH Key  
 FT exon  
 1942..2383  
 /tag- a  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

2384..2443  
 /tag- b  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

2444..2597  
 /tag- c  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

2598..2677  
 /tag- d  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

2678..3012  
 /tag- e  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

exon  
 2678..3012  
 /tag- e  
 /note- "these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"

```

FT      /tag- e
PN      WO9201792-A.
PD      06-FEB-1992.
PF      24-JUL-1991; F00607.
PR      24-JUL-1990; FR-009460.
PA      (SNFI) SANOFI SA.
PI      (ERAP) SOC NAT ELF AQUITAINE.
PI      Dubois M, Grlson R, Leguay JJ, Pignard A, Toppan A;
DR      WPI: 92-064950/08.
DR      P-PSDB: R20820.
PT      New recombinant protein with endo chitinase activity - also DNA
PT      encoding it and plants which express it, resistant to fungi,
PT      insects, bacteria and nematodes.
PS      Example: Fig 2; 82pp; French.
CC      The inventors claim a new recombinant gene which codes for a protein
CC      with encochitinase activity, or its precursor (see Q21007). The
CC      coding part of the gene contains at least the 3' part of DNA for
CC      tomato endochitinase and at least the 3' part of DNA for the tobacco
CC      enzyme (DNA can be genomic or cDNA, but at least one intron is pref.
CC      present). The recombinant gene includes the 35S promoter of
CC      cauliflower mosaic virus plus the terminator from the nopaline
CC      synthase gene of Agrobacterium tumefaciens.
SQ      Sequence 3012 BP; 965 A; 634 C; 436 G; 977 T;

```

```

Query Match
Best Local Similarity 48.9%; Score 34; DB 1; Length 3012;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```

```

OY      10 CTCCTCTCTGATTAATTCCTTCACTGCAATTGCAAGATTACATTACT 69
DB      405 CCCCCCCCCCAAAAAAATTTAAATTTGTTTAAAAAATTTTCAATTCAAA 464
OY      70 GTGATGATATGTGTGCAAAAAAAGGCTTTGTTAAATTAATTTACTGTTGT 129
DB      465 ATTTTATTTTCAATTTAGTAAATAAGATATTCGAAAAACATTTTACTATT 524
OY      130 GAATCCATGTTGCTTTTCCCATGACTAGTACATTAACCATCTGCACTGTAGA 189
DB      525 CAAAAACCAACACTTTTCCGAAAAAATTTCTATTACACCAACAATATGAGAAAT 584
OY      190 AAAACA 195
DB      585 AATCA 590

```

RESULT 7

003253/c

ID 003253 standard; DNA: 4483 BP.

AC 003253;

DT 25-JUL-1990 (first entry)

DE Sequence of clone pBSgluc3.1 encoding beta-1,3-glucanase from genomic

DE sequence.

KW Beta-1,3-glucanase; papermaking; polysaccharide hydrolase;

KW ds.

OS Nicotiana sp.

PN EP-353191-A.

PD 31-JAN-1990.

PF 20-JUL-1988; 810555.

PR 29-JUL-1988; US-353112.

PA (CIBA) Ciba Geigy AG.

PI Shinsui H, Wenzler H, Hofsteenge J, Ryals J, Sperisen C;

DR WPI: 90-031717/05.

PT Recombinant DNA encoding beta-1,3-glucanase polypeptide(s) -

PT useful in plant protection and biomass conversion pref. isolated

PT from tobacco.

PS Disclosure; Fig 8; 41pp; English.

CC Beta-1,3-glucanase is involved in plant defence against pathogens and as

CC it hydrolyses polysaccharides, useful in biomass conversion or paper

CC making. The clones allow for transfer of the gene to other species and

CC control of gene expression.

SQ Sequence 4483 BP; 1418 A; 838 C; 744 G; 1483 T;

```

Query Match
Best Local Similarity 52.1%; Score 33.6; DB 1; Length 4483;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

```

OY      1 GCTCTTCTCCCTCCTCTGAAATTAATTCCTTCACTGCAATTGCAAGATTACA 60
DB      4481 GCTCTTCTGCTGCTCTGCAATAGAGCTCTTCTTCCCTTGGAGTGAATGGA 4422
OY      61 CATTTCACGTGATGATATTTGTGTCAAAAAAGAGTCTTTGTTAAATTAAC 120
DB      4421 CAATTCACACGAGCTAGTTGTGGTAAGATCAAAATGATCTTTAGCTGGAATTA 4362
OY      121 TTGTTTGAATCCATCTGCTT 144
DB      4361 CCGGTTTGTGGCATGCTTCTT 4338

```

RESULT 8

```

ID N90843/c
AC N90843;
DE 24-JAN-1990 (first entry)
DE Tobacco basic beta-1,3 glucanase gene.
KW Tobacco basic beta-1,3 glucanase; pBSgluc3.3; herbicide resistance.
OS Nicotiana tabacum cv. Xanthi.
PN EP-332104-A.
PD 13-SEP-1989.
PF 06-MAR-1989; 103888.
PR 08-MAR-1988; US-165667.
PA (CIBA) Ciba Geigy AG.
PI Ryals J, Montoya A, Harms C, Duesing J, Sperisen C, Melus F,
PI Payne G;
DR WPI: 89-265342/37.
PT Chemical induction of cloned genes in plants - DNA sequence and vectors
PT and processes for inducible expression leading to herbicide and insect
PT resistance.
PS Claim 79; page 72-74; 117pp; English.
CC Sequence is a chemically inducible gene used to regulate
CC transcription of an associated DNA sequence in plant
CC tissue.
CC See also N90368, N90841-N90846, and P91384-P91386.
SQ Sequence 4483 BP; 1418 A; 837 C; 745 G; 1483 T;

```

```

Query Match
Best Local Similarity 52.1%; Score 33.6; DB 1; Length 4483;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

```

OY      1 GCTCTTCTCCCTCCTCTGAAATTAATTCCTTCACTGCAATTGCAAGATTACA 60
DB      4481 GCTCTTCTGCTGCTCTGCAATAGAGCTCTTCTTCCCTTGGAGTGAATGGA 4422
OY      61 CATTTCACGTGATGATATTTGTGTCAAAAAAGAGTCTTTGTTAAATTAAC 120
DB      4421 CAATTCACACGAGCTAGTTGTGGTAAGATCAAAATGATCTTTAGCTGGAATTA 4362
OY      121 TTGTTTGAATCCATCTGCTT 144
DB      4361 CCGGTTTGTGGCATGCTTCTT 4338

```

RESULT 9

```

ID V72993/c
AC V72993 standard; cDNA: 4483 BP.
AC V72993;
DE 25-FEB-1999 (first entry)
DE Tobacco basic beta-1,3-glucanase genomic DNA clone pBS-gluc3.1.
KW Regulation: transcription; plant tissue; chimeric construction; PT:
KW pathogenesis-related protein; anti-pathogenic; transgenic plant;
KW beta-1,3-glucanase activity; pest resistance; ss.
OS Nicotiana sp.
PN US5847258-A.

```



FD 08-DEC-1998 .  
PF 31-MAY-1995; 457364 .  
PR 31-MAY-1995; US-457264 .  
PR 08-MAR-1988; US-163567 .  
PR 06-FEB-1989; US-305566 .  
PR 24-MAR-1989; US-329018 .  
PR 20-JUN-1989; US-358672 .  
PR 20-OCT-1989; US-425504 .  
PR 07-SEP-1990; US-580431 .  
PR 21-DEC-1990; US-632441 .  
PR 01-APR-1991; US-678378 .  
PR 27-SEP-1991; US-768122 .  
PR 06-MAR-1992; US-848506 .  
PR 06-NOV-1992; US-973197 .  
PR 06-APR-1993; US-042847 .  
PR 12-APR-1993; US-045957 .  
PR 16-JUL-1993; US-093301 .  
PR 13-JAN-1994; US-181271 .  
PA (NOVS ) NOVARTIS FINANCE CORP .  
PI Moyer MB, Payne GB, Ryals JA, Ward ER;  
PI WPI; 99-059180/05 .  
PT DNA encoding pathogenesis-related glucanase proteins - useful for  
PT producing transgenic plants with enhanced disease or pest resistance  
PS Example 33: column 163-168; 169pp; English .  
PS The present invention describes a DNA molecule encoding a  
PS pathogenesis-related (PR) protein having beta-1,3-glucanase activity  
CC selected from PR-2, PR-2', PR-2'', PR-N, PR-O and PR-O' . Also described  
CC are: (i) a chimeric gene comprising the above DNA molecule linked to a  
CC heterologous promoter; (ii) a vector containing the chimeric gene;  
CC (iii) a host cell containing the chimeric gene; (iv) a transgenic plant  
CC containing the chimeric gene; and (v) a seed from the transgenic plant .  
CC The DNA molecule is used to produce transgenic plants with enhanced  
CC disease or pest resistance . The present sequence represents a tobacco  
CC basic beta-1,3-glucanase genomic DNA sequence .  
SO Sequence 4483 BP; 1418 A; 838 C; 744 G; 1483 T;

Query Match	Similarity	9.2%	Score 33.6	DB 1	Length 4483
Best Local	Similarity 52.1%		Pred. No. 3.4		
Matches 75	Conservative	0	Mismatches 69	Indels	Gaps 0
Oy	1	GCTCTTTCCTCCTCCCTCCATTAATTTCTTCACACTGCAATTTGCAAGATTACA	60		
Db	4481	GCTCTTTCCTCCTCCCTCCATTAAGAGTCTTCTTCTGCGCTTTGGAGCTGAGAAATGCA	4422		
Oy	61	CATTTCACCTGTCATGTATATTTGTGTGCACAAAAAAGTGTCTTTGTTAAATTAC	120		
Db	4421	CATTTCACACGAGGTAGCTTTGTGGGTAAAGATCAAAATGATCTCTTACCTGGAAGTTAT	4362		
Oy	121	TTGTTTGGAAATCCATCTTGCTT	144		
Db	4361	CCGCTTTGTGTGGCATGCTTCTT	4338		
RESULT 10					
V62789/c					
ID	V62789	standard; cDNA; 4483 BP.			
AC	V62789				
DT	05-MAR-1999	(first entry)			
DE	Tobacco beta-1,3-glucanase gene clone PRSGluc39.1.				
KW	Chemically regulatable DNA promoter; expression control; pesticide;				
OS	herbicidal tolerance; beta-1,3-glucanase gene; ss.				
PN	Nicotiana glauca.				
PD	US5851766-A.				
PF	22-DEC-1998.				
PR	31-MAY-1995; 456262.				
PA	31-MAY-1995; US-456262.				
PI	(NOVS) NOVARTIS FINANCE CORP.				
DR	Harms C, Ryals JA.				
WP	99-080396/07.				
PT	Isolating chemically regulatable DNA sequences in plants - useful				
PS	for chemically controlling expression in transformed plants				
Example 33	Column 169-174; 175pp; English.				

CC This sequence represents a clone of the tobacco beta-1,3-glucanase gene.  
CC This gene can be isolated using the method of the invention.  
CC The method is for isolating a chemically regulatable DNA promoter  
CC fragment from the 5' flanking region of a chemically regulatable gene in  
CC a plant tissue. The method allows isolation of sequences which will be  
CC useful for the controlled expression of genes, under the control of a  
CC non-coding regulatable sequence. This is useful in plants with a  
CC herbicide or pesticide detoxification mechanism under the control of a  
CC chemical regulator, the regulator being applied before or with the  
CC herbicide or pesticide to give optimal tolerance. The promoter fragment  
CC is useful for controlling sequences which encode traits such as  
CC height, shape, development, male or female sterility, and the ability  
CC of the plant to withstand cold, heat, salt and drought. The chemical  
CC induction of the promoter allows the regulation of production of  
CC compounds, e.g. flavours, fragrances, pigments, natural sweeteners,  
CC industrial feedstocks, antimicrobials and pharmaceuticals, by  
CC biosynthesis or metabolite conversion, whose biosynthesis is controlled  
CC by endogenous or foreign genes. The method allows control over the time  
CC and rate of gene expression either throughout the whole plant, or in  
CC localized tissues, to achieve e.g. fungal or insect resistance by for  
CC instance dusting the leaves with the chemical regulator. Controlling the  
CC developmental processes by the application of a regulating chemical in  
CC e.g. the commercial production of cultivated crops allows processes such  
CC as germination, flower formation and fruit ripening to be synchronised at  
CC a given time.  
CC Sequence 4483 BP; 1418 A; 838 C; 744 G; 1483 T;

```

Query Match          9.2%; Score 33.6; DB 1; Length 4483;
Best Local Similarity 52.1%; Pred. No. 3.4;
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0

OY      1 GCCTTCTCTCCTCCCTCCTCTGATTAATTAACTTCATCACTGCAGTTGCAAGATTACA 60
        ||||| | | | | |
Db       4481 GCCTTTTGCTGGCTCTCTGAATAGAGTCTTCTTCCTCGCTTTGGAGTGTGATAAATGGA 4422
OY      61 CATTCACGTGATGATATGTTGTTGCCAAAAAAAAGAAGCTTGTGTTAAATATAC 120
        || ||||| | | | | | | | | | | | | | | | | | | | | |
Db       4421 CAATTCACAGCGACTAGTTTGGGGTAAAGATCAAATAAGATCTTTAGCTGGAAGTTAT 4362
OY      121 TTGGTTGATCATCATCTTGCCTT 144
        ||||| | | | | |
Db       4361 CCGCTTGTGTGTCAGATGCTTCTT 4338

RESULT  11
V74317/c
ID       V74317 standard; DNA: 3601 BP.
AC       V74317;
DE       16-MAR-1999 (first entry)
DT       Staphylococcus aureus contig SEQ ID #6.
KW       Computer readable medium; vaccine; S.aureus infection; Immunodetection;
          KRN cellulitis; eyelid infection; food poisoning; Osteomyelitis; therapy;
          skin infection; surgical wound infection; scalded skin syndrome;
          toxic shock syndrome; ds.
KM       Staphylococcus aureus.
OS       Staphylococcus aureus.
FH       key location/Qualifiers
FT       misc-feature 601..660
          /tag=a
          /note="these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT       misc-feature 2401..2460
          /tag=b
          /note="these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FN       EP-786519-A2.
FD       30-JUL-1997.
FE       07-JAN-1997; 100117.
FR       05-JAN-1996; US-009861.

```

PA (HUNA-) HUMAN GENOME SCI INC.  
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,  
PI Rosen CA:  
DR WPI 97-374922/35.  
PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -  
PT stored on computer readable medium and used in the production of  
PS anti-*S.aureus* vaccines  
PS Claim 1: Page 238-240: 3271pp: English.  
CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences  
CC of the invention. The DNA sequences are recorded on a computer readable  
CC medium, preferably selected from a floppy or hard disk, random access  
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using  
CC the *S.aureus* DNA sequences allows putative functions to be assigned so  
CC that protein-encoding or regulatory regions of commercial, therapeutic or  
CC industrial importance can be obtained. Specifically, sequences which are  
CC likely to encode antigens have been identified and these polypeptides can  
CC be used in a vaccine composition against *S.aureus* infection. The  
CC polypeptides can also be used in a kit for the immunodetection of  
CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,  
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,  
CC skin and surgical wound infections, scalded skin syndrome, toxic shock  
CC syndrome, etc. Organisms transformed with the DNA sequences can be used  
CC for recombinant production of the polypeptides. The new DNA sequences  
CC (and their fragments) are useful as primers or probes for isolating  
CC homologues of any of the *S.aureus* DNA sequences contained on the  
CC computer readable medium.  
SQ Sequence 3601 BP: 1170 A: 496 C: 738 G: 1057 T:

Query Match 9.1%; Score 33.4; DB 1; Length 3601;  
Best Local Similarity 50.3%; Pred. No. 3.6;  
Matches 82: Conservative 0; Mismatches 81: Indels 0; Gaps 0;

OY 14 CCCCCCTGAATTATTTCTTCAACTTGCATTTGCAGGATTACATTTACTGTGA 73

DB 3577 CGACGACCTCTTATTTTGGCACGATCCCTTCCAACTTGTGCCCGTA 3518

OY 74 TGTATATGTGTGCAAAAAAAGTCTTGTAAATTAATCTGTTGTTGAAT 133

DB 3517 TTTTTCAGTGTGTCATCATCATCATCATCATCATCATCATCATCATTAAGTT 3458

OY 134 CCATCTTGTCTTCCCATTTGGCACTAGTCAATTAACCATCT 176

DB 3457 AATAATACATTTTTCATCATCTGTAAGATCATTAAGTATCT 3415

RESULT 12

X20248\_08/c  
Continuation (9 of 10) of X20248 from base 800001 (Borrelia burgdorferi polynucleotide S

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 8.8%; Score 32.2; DB 1; Length 110000;  
Best Local Similarity 53.6%; Pred. No. 19;  
Matches 67: Conservative 0; Mismatches 58: Indels 0; Gaps 0;

OY 205 CTAGCTATCAGCATCTGACAGTGAATGGATGCTCAGACATTTCCACCGAGACA 264

DB 103178 CAACGTATATTCATATAAACAAATTTTGGCATTTTCATCATTTCCATGATACATGCTGA 103119

OY 265 GCCTGTTCTATCCGTGTTAATAATTAGTTGGTCTCTACATCATTAACCAACCTG 324

DB 103118 AAGTATTTCTACATGTTTCAAAATATTTATTCAGATTTTCAAAATGCCGAGCATGCTGA 103059  
OY 325 CTCGA 329.  
DB 103058 ATCCA 103054

RESULT 13

X20248\_09/c  
Continuation (10 of 10) of X20248 from base 900001 (Borrelia burgdorferi polynucleotide

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP	Fragment Name	Begin	End
WP	X20248_00	1	110000
WP	X20248_01	100001	210000
WP	X20248_02	200001	310000
WP	X20248_03	300001	410000
WP	X20248_04	400001	510000
WP	X20248_05	500001	610000
WP	X20248_06	600001	710000
WP	X20248_07	700001	810000
WP	X20248_08	800001	910000
WP	X20248_09	900001	910715

Query Match 8.8%; Score 32.2; DB 1; Length 10715;  
Best Local Similarity 53.6%; Pred. No. 10;  
Matches 67: Conservative 0; Mismatches 58: Indels 0; Gaps 0;

OY 205 CTAGCTATCAGCATCTGACAGTGAATGGATGCTCTCAGAACCATTTCCACCGAGACA 264

DB 3178 CAACGTATATTCATATAAACAAATTTTGGCATTTTCATCATTTCCATGATACATGCTGA 3119

OY 265 GCCTGTTCTATCCGTGTTAATAATTAGTTGGTCTCTACATCATTAACCAACCTG 324

DB 3118 AAGTATTTCTACAAAGTTTCAAAATATTTATTCAGATTTTCAAAATGCCGAGCATGCTGA 3059

OY 325 CTCGA 329

DB 3058 ATCCA 3054

RESULT 14

V02032/c  
ID V02032 standard; DNA: 40352 BP.

AC 12-JUN-1998 (first entry)

DE MAGE-B cluster DNA sequence.

KW MAGE-B cluster DNA; testicular seminoma; cancer; screening; leukaemia;

OS carcinoma; ss.

OS Homo sapiens.

FN key Location/Qualifiers

FT prim\_transcript 3266..7979

FT /tag a

FT /label B2 gene

FT /tag b

FT /note "exon 1 of B2 gene"

FT /tag c

FT /note "intron 1 of B2 gene"

FT /tag d

FT /note "exon 2 of B2 gene"

FT /tag e

FT /note "coding sequence for B2 gene"

FT /tag f

FT /note "poly-A signal for B2 gene"

FT /tag g

FT /note "exon 1 of B3 gene"

FT CDS 23607..24647

FT /note- h  
FT /tag- "coding sequence for B3 gene"  
FT 25152 .25157  
FT polyA\_signal  
FT /tag- 1  
FT /note- "poly-A signal for B3 gene"  
FT prim\_transcript 29748 .31827  
FT /tag- j  
FT /label- B4 gene  
FT CDS 29808 .30848  
FT /tag- k  
FT /note- "coding sequence for B4 gene"  
FT 31822 .31827  
FT /tag- l  
FT /note- "poly-A signal for B4 gene"  
FT prim\_transcript 31403 .39691  
FT /tag- m  
FT /label- B1 gene  
FT exon 31403 .31474  
FT /tag- n  
FT /note- "exon 1 of B1 gene"  
FT intron 31475 .33957  
FT /tag- o  
FT /note- "intron 1 of B1 gene"  
FT exon 33958 .34062  
FT /tag- p  
FT /note- "exon 2 of B1 gene"  
FT intron 34063 .35056  
FT /tag- q  
FT /note- "intron 2 of B1 gene"  
FT exon 35057 .35139  
FT /tag- r  
FT /note- "exon 3 of B1 gene"  
FT intron 35140 .38087  
FT /tag- s  
FT /note- "intron 3 of B1 gene"  
FT exon 38088 .39691  
FT /tag- t  
FT /note- "exon 4 of B1 gene"  
FT CDS 38148 .39191  
FT /tag- u  
FT /note- "coding sequence for B1 gene"  
FT polyA\_signal 39674 .39679  
FT /tag- v  
FT /note- "poly-A signal for B1 gene"  
FT W09746710-A1.  
PD 11-DEC-1997.  
PF 05-JUN-1997: U09774.  
PR 25-APR-1997: US-846111.  
PR 05-JUN-1996: US-658578.  
PA (LUDW-) LUDWIG INST CANCER RES.  
PI Boon T, Brasseur F, Lurquin C;  
DR WPI: 98-042216/04.  
PT Screening for the possibility of testicular seminoma and other  
PS Claim 7: Pages 22-34: 40pp: English.  
CC This is the MAGE-B cluster DNA sequence. The invention provides a novel  
CC method for screening for the possibility of a testicular seminoma,  
CC non-small cell lung carcinoma, melanoma, breast cancer, sarcoma, or  
CC leukemia in a sample. The method comprises contacting the sample with  
CC at least one nucleic acid molecule which hybridizes to mRNA corresponding  
CC to an MAGE-Xp gene, and determining hybridization as a determination of  
CC possible presence of testicular seminoma, non-small cell lung carcinoma  
CC melanoma, breast cancer, sarcoma, or leukemia in the sample. By assaying  
CC for the MAGE-B1 gene, or MAGE-B2 gene, both contained within this 40352  
CC base pair MAGE-B cluster DNA sequence, the presence of such cancers in a  
CC sample can be determined. The genomic DNA that encodes the MAGE-B2 gene  
CC consists of nucleotides 3266-7979 of this MAGE-B cluster DNA sequence.  
CC The nucleic acid molecule that encodes a MAGE-B1 gene, consists of, in  
CC 5' to 3' order, nucleotides 31403-31474, 33958-34062, 35057-35139, and  
CC 38088-39691, nucleotides 31403-31474, 33958-34062, and 38088-39691,  
CC and nucleotides 35057-35139 and 38088-39691, and nucleotides 33958-34062 and  
CC 38088-39691 of this sequence. The primers specific for the MAGE-B  
CC gene can be used in a kit to amplify a MAGE-B gene.

SO Sequence 40352 BP; 11599 A; 8654 C; 9772 G; 10327 T;  
Query Match 8.7%; Score 32; DB 1; Length 40352;  
Best Local Similarity 54.2%; Pred. No. 17;  
Matches 65; Conservative 0; Mismatches 55; Indels 0; Gaps 0;  
QY 32 CTTTCACCTGGCAATTCGCAAGATGATACATTCCTGATGATATGTCGCAAA 91  
DB 19244 CTTGCTCCTCCCAAAATGGGATTCAGCGGACCCAGACCCCTTAATAATTA 19185  
QY 92 AAAAAAAGTGTCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 151  
DB 19184 AAAAAAAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19125  
RESULT 15  
X15658  
ID X15658 standard; DNA; 1407 BP.  
AC X15658:  
DT 07-MAY-1999 (first entry)  
DE Protein phosphatase gene coding sequence of Saccharomyces cerevisiae.  
KW Protein phosphatase gene; growth; fermentation activity;  
KM dough production; yeast; ds.  
OS Saccharomyces cerevisiae.  
PN J11042090-A.  
PD 16-FEB-1999.  
PE 29-JUL-1997: 203652.  
PR 29-JUL-1997: JP-203652.  
PA (KANF ) KANEKA CORP.  
PA (SHOS ) SHOMA SANGYO CO.  
DR WPI: 99-197822/17.  
PT New yeast of controlled activation at low temperatures - useful for  
PT improving the quality of dough  
PS Claim 5; Page 21-22; 41pp: Japanese.  
CC The present sequence represents a protein phosphatase gene sequence  
CC of Saccharomyces cerevisiae. The specification describes new  
CC S. cerevisiae in which the growth and/or the fermentation activity  
CC is controlled at least in the range of 0-20 degrees Celsius. These  
CC yeast are prepared by deleting the function of at least one protein  
CC phosphatase gene. The yeast is useful in the production of dough.  
SQ Sequence 1407 BP; 434 A; 222 C; 325 G; 426 T;  
Query Match 8.7%; Score 32; DB 1; Length 1407;  
Best Local Similarity 46.8%; Pred. No. 6; Gaps 0;  
Matches 101; Conservative 0; Mismatches 115; Indels 0; Gaps 0;  
QY 21 TGAATTAATTCCTTCAACTGCAATTCGACAGATTAACATTCCTGATGATAT 80  
DB 891 TGAATCGAGTCACACATGTTGAGCGTATGATCAAAAAATTAATTAATTCACAAAGTC 950  
QY 81 TGTGTCCAAAAAAAGAGTCTTGTAAATTAATTAATTAATTAATTAATTAATTAATTA 140  
DB 951 TTTTGTCCCAAGAGGAAGATATTTTGAATTCATGATTTTGGATGATGATGA 1010  
QY 141 GCTTTTCCCATTTGAACTAGTCAATTAACCATCTCTGAACTGGTAGAAAAACATTTGA 200  
DB 1011 AGTGTTCGCAATTAACCAAAAAAATTAACAAAGCCCTTGATGATGATGAAGATATGA 1070  
QY 201 AGACCTAGTCAATGACATCGAAGTGAATTGA 236  
DB 1071 CGACATGAATTAATGATGATTTGATACCAATTGA 1106

Search completed: September 25, 1999, 09:56:03  
Job time: 8014 sec





[illegible]

```

RESULT 2
US-08-475-427-5
: Sequence 5, Application US/08475427
: Patent No. 5859340
: GENERAL INFORMATION:
: APPLICANT: DUBOIS, Michel
: APPLICANT: GRISON, Rene
: APPLICANT: LEGUAY, Jean-Jacques
: APPLICANT: PIGNARD, Annie
: APPLICANT: TOPPAN, Alain
: TITLE OF INVENTION: Recombinant gene coding for a protein
: TITLE OF INVENTION: having endonuclease activity or for a precursor thereof
: NUMBER OF SEQUENCES: 15
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 3000 K Street, N.W., Suite 500
: CITY: Washington, D.C.
: COUNTRY: USA
: ZIP: 20007-5109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/475.427
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/842,165
: FILING DATE: 01-MAY-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/FR91/00607
: FILING DATE: 21-JUL-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 90 09460
: FILING DATE: 24-JUL-1990
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 16781/564/BDL
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
: TELEX: 904136
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3012 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: intron
: LOCATION# 2384..2462

```

```

? OTHER INFORMATION: /number= 1
? FEATURE:
? NAME/KEY: inttron
? LOCATION: 2617..2697
? OTHER INFORMATION: /number= 2
? FEATURE:
? NAME/KEY: CDS
? LOCATION: join(1942..2383, 2463..2616, 2698..3007)
? FEATURE:
? NAME/KEY: sig_peptide
? LOCATION: 1942..2166
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: join(2167..2383, 2463..2616, 2698..3007)
? LOCATION: 2658)
US-08-475-427-5

```

Query Match	9.38;	Score 34;	DB 4;	Length 3012;
Best Local Similarity	48.98;	Pred. NO. 0.74;		
Matches	91;	Conservative	0;	Mismatches 95;
			Indels	0;
			Gaps	0;

QY	10	GTCCTCCTCGATATTAAATCTTCAACTGCAATTTGGAAGATTACCATTTACT	69
Dd	405	CCCCCACCCCAAAAAAATATTTAAATTTGTTTAAAAAATTTTCAATTTCAA	468
QY	70	GTGATGTAATATGTGTCGCAAAAAAAAAGTGCTTTGTATAAATTACTGGTTGT	129
Dd	465	ATTATTTATTTCTATCTAGTAAATAATATATATCTCAAAAAACATTTTTTCTATT	528
QY	130	GAATCGATCTTGCTTTTCCCATTGGACAAGTATTAACCATCTCTGAACTGGTAGA	189
Dd	535	CAAAAACCAACACACTCTTTCCAGAAAAAAATTTCTATTCACCAACCAAAATATGAGAAAT	594
QY	190	AAAAA 195	
Dd	585	AAATCA 590	

RESULT 3  
 : Sequence 5, Application US/08181271A  
 : Patent No. 5614395  
 : GENERAL INFORMATION:  
 APPLICANT: Ryals, John A.  
 APPLICANT: Alexander, Danny C.  
 APPLICANT: Beck, James J.  
 APPLICANT: Duesing, John H.  
 APPLICANT: Friedrich, Leslie B.  
 APPLICANT: Goodman, Robert M.  
 APPLICANT: Harms, Christian  
 APPLICANT: Meins, Jr., Frederick  
 APPLICANT: Montoya, Alice  
 APPLICANT: Moyer, Mary B.  
 APPLICANT: Neuhaus, Jean-Marc  
 APPLICANT: Payne, George B.  
 APPLICANT: Sperison, Christoph  
 APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Uknes, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Shericea C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10533  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,271A
FILING DATE: 13-JAN-94
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/GGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4483 base pairs.
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-181-271A-5

```

```

0Y          61  CATTTCAGTGGATGATGATGTTGGTGCACAAAAAAGTCTTTGTTAAATAC 120
              || ||||| || || ||||| || ||||| || || ||||| || || |||||
Db          4421  CAAATTCACACGGAGCTTGTTGGGTGAATCAACAAATGATCTTACTGGAATTAT 4362
              ||||| || || |||||
0Y          121  TTGGTTGTGTAATCCATCTTGGCTT 144
              ||||| || || |||||
Db          4361  CCGCTTGTGTGGCATGCTTCTT 4338
              ||||| || || |||||

RESULT      4
US-08-449-315-5/C
: Sequence 5, Application US/08449315
: Patent No. 5650505
: GENERAL INFORMATION:
: APPLICANT: Ryals, John A.
: APPLICANT: Alexander, Danny C.
: APPLICANT: Beck, James J.
: APPLICANT: Duesing, John H.
: APPLICANT: Friedrich, Leslie B.
: APPLICANT: Goodman, Robert M.
: APPLICANT: Harms, Christian
: APPLICANT: Weins, Jr., Frederick
: APPLICANT: Montoya, Alice
: APPLICANT: Moyer, Mary B.
: APPLICANT: Neuhaus, Jean-Marc
: APPLICANT: Payne, George B.
: APPLICANT: Sperison, Christoph
: APPLICANT: Stinson, Jeffrey R.
: APPLICANT: Uknes, Scott J.
: APPLICANT: Ward, Eric R.
: APPLICANT: Williams, Sherlica C.
: TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
: TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
: NUMBER OF SEQUENCES: 106
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CIBA-GEIGY Corporation.
: STREET: 7 Skyline Drive
: CITY: Hawthorne
: STATE: New York
: COUNTRY: USA
: ZIP: 10532
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/449,315
: FILING DATE: 24-MAY-1995
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/181,271
: FILING DATE: 13-JAN-94
: APPLICATION NUMBER: US 08/093,301
: FILING DATE: 16-JUL-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/937,197
: FILING DATE: 6-NOV-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/678,378
: FILING DATE: 1-APR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/305,566
: FILING DATE: 6-FEB-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/165,667
: FILING DATE: 8-MAR-1988
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/042,847
: FILING DATE: 6-APR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/632,441

```





TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-444-803-5

Query Match 9.2%; Score 33.6; DB 2; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GCCTTCTCTCCCTCCTGATTAATTCCTTCACTTGCATTTGCAAGATTACA 60  
DB 4481 GCCTTTTGTGCTGCTCTGATAGAGCTTCTTCTGCTTGGAGCTGATGAAATGA 4422  
QY 61 CATTTCACTGATGATATATTGTGTCAGAAAAAAGTCTTGTAAATTAC 120  
DB 4421 CAATTCACAGGAGCTGATTGTGGTAAGATCAAAATAGATCTTAGCTGGAATTAT 4362  
QY 121 TTGTTTGTGAATCCATCTTGCTT 144  
DB 4361 CCGTTTGTGTGCGATGCTTCTT 4338

## RESULT 6

US-08-449-043-5/C  
Sequence 5, Application US/08449043  
Patent No. 5689044

## GENERAL INFORMATION:

APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedlich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/449,043  
FILING DATE: 24-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94

APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/PL/CCG 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-449-043-5

Query Match 9.2%; Score 33.6; DB 2; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 1 GCCTTCTCTCCCTCCTGATTAATTCCTTCACTTGCATTTGCAAGATTACA 60  
DB 4481 GCCTTTTGTGCTGCTCTGATAGAGCTTCTTCTGCTTGGAGCTGATGAAATGA 4422  
QY 61 CATTTCACTGATGATATATTGTGTCAGAAAAAAGTCTTGTAAATTAC 120  
DB 4421 CAATTCACAGGAGCTGATTGTGGTAAGATCAAAATAGATCTTAGCTGGAATTAT 4362  
QY 121 TTGTTTGTGAATCCATCTTGCTT 144  
DB 4361 CCGTTTGTGTGCGATGCTTCTT 4338

RESULT 7  
US-08-456-265A-5/C  
Sequence 5, Application US/08456265A  
Patent No. 5767369  
GENERAL INFORMATION:  
APPLICANT: Alexander, Danny C.  
APPLICANT: Ryals, John A.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Stinson, Jeffrey R.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF SEQUENCES: 111  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 520 White Plains Road, P.O. Box 2005  
CITY: Tarrytown  
STATE: New York  
COUNTRY: USA  
ZIP: 10591  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,265A  
FILING DATE: 31-MAY-95  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/181,271  
FILING DATE: 13-JAN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-19825/P1/C/GC 1727/DIV10  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8587  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-456-265A-5  
Query Match 9.2%; Score 33.6; DB 3; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 1 GCTCTTCTCTCCCTCCTCGAATTATCTTCACTTGCAATTTGCAAGGATTACA 60  
DB 4481 GCTCTTGTGCTCTCTCGAATAGAGTCTTCTTCCCTTGAGGATTGAATGGA 4422  
QY 61 CATTTCACGTGATGTATATGTGTGCAAAAAAAGTGTCTTTTAATATAC 120  
DB 4421 CATTTCACGAGGAGCTAGTGTGGTAAGATCAAAATGATCTTACGTGAAGTTAT 4362  
QY 121 TTGGTTTGATCCATCTGCTT 144  
DB 4361 CCCGTTGTGTGCGATCTCTT 4338  
RESULT 8  
US-08-455-416-5/C  
Sequence 5, Application US/08455416  
Patent No. 5777200  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedrich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Uknes, Scott J.  
APPLICANT: Ward, Eric R.  
APPLICANT: Williams, Shericea C.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,416  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/PL/GCC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-455-416-5

Query Match 9.28; Score 33.6; DB 3; Length 4483;  
Best Local Similarity 52.18; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

DB 4481 GCCTTTCTCTCCCTCTCGATTTAATCTTTCAACTTGCATTTGCAAGATTACA 60  
1 GCTTTCTCTCTCCCTCTCGATTTAATCTTTCAACTTGCATTTGCAAGATTACA 60  
4481 GCCTTTCTCTCTCGATTTAATCTTTCTTCTTCCCTTTGAGATTGATGAATGGA 4422

QY 61 CATTTCAGTGTATATGTTGCAAAAAAAAAAGTCTTTTAATAATTAC 120  
DB 4421 CAATTCACGAGAGCTAGTTGTGGGTAAGATCAAAATAGATCTTACGTGAAGTTAT 4362  
QY 121 TTGTTTGTCAATCATCTTCTT 144  
DB 4361 CCGGTTTGTGTGGCAGCTCTT 4338

RESULT 9  
US-08-455-244-5/C  
Sequence 5, Application US/08455244  
Patent No. 5789214  
GENERAL INFORMATION:  
APPLICANT: Ryals, John A.  
APPLICANT: Alexander, Danny C.  
APPLICANT: Beck, James J.  
APPLICANT: Duesing, John H.  
APPLICANT: Friedlich, Leslie B.  
APPLICANT: Goodman, Robert M.  
APPLICANT: Harms, Christian  
APPLICANT: Meins, Jr., Frederick  
APPLICANT: Montoya, Alice  
APPLICANT: Moyer, Mary B.  
APPLICANT: Neuhaus, Jean-Marc  
APPLICANT: Payne, George B.  
APPLICANT: Sperison, Christoph  
APPLICANT: Stinson, Jeffrey R.  
APPLICANT: Utnes, Scott J.  
APPLICANT: Ward, Eric R.  
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: CIBA-GEIGY Corporation  
STREET: 7 Skyline Drive  
CITY: Hawthorne  
STATE: New York  
COUNTRY: USA  
ZIP: 10532  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,244  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER:  US 07/425,504
3      FILING DATE:  20-OCT-1989
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER:  US 07/848,506
6      FILING DATE:  6-MAR-1992
7      PRIOR APPLICATION DATA:
8      APPLICATION NUMBER:  US 07/768,122
9      FILING DATE:  27-SEP-1991
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER:  US 07/560,431
12     FILING DATE:  7-SEP-1990
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER:  US 07/368,672
15     FILING DATE:  20-JUN-1989
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER:  US 07/329,018
18     FILING DATE:  24-MAR-1989
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER:  US 08/045,957
21     FILING DATE:  12-APR-1993
22     ATTORNEY/AGENT INFORMATION:
23     NAME:  Elmer, James Scott
24     REGISTRATION NUMBER:  36,129
25     REFERENCE/DOCKET NUMBER:  S-19825/P1/CGC 1722
26     TELECOMMUNICATION INFORMATION:
27     TELEPHONE:  (919)541-8614
28     TELEFAX:  (919)541-8689
29     INFORMATION FOR SEQ ID NO:  5:
30     SEQUENCE CHARACTERISTICS:
31     LENGTH:  4483 base pairs
32     TYPE:  nucleic acid
33     STRANDEDNESS:  single
34     TOPOLOGY:  linear
35     MOLECULE TYPE:  CDNA
36     OS-08-455-244--5

```

Query Match	9.28;	Score 33.6;	DB 3;	Length 4483;
Best Local Similarity	52.18;	Pred. No. 1.1;		
Matches 75;	Conservative 0;	Mismatches 69;	Indels 0;	Gaps 0

Oy	1	GCTCTTCTCCTCCCTCCCTCGATTAATTCCTTCACTGCAATTGCAGAGTTACA	60
Db	4481	GCTCTTTTGGCTCTCTTGAAATAGAGCTTCTTCCCTGCTTTGAGTTGAATGAATGA	44222
Oy	61	CATTTCAGTGATGATATATTGTGTGCAGAAAAAAAAGTGTCTTTTAAATTAC	120
Db	4421	CAATTACACGAGACTGTTGTGGGTAAAGATCAAAATAGATCTTTACTGGAAGTTAT	4362
Oy	121	TTGGTTGTGAATCCATCTTGCTT	144
Db	4361	CCCGTTTGTGTGGCATGCTTCTT	4338

US-RSULT 10  
: Sequence 5, Application US/08454876  
: Patent No. 5804693  
: GENERAL INFORMATION:  
: APPLICANT: Ryals, John A.  
: APPLICANT: Alexander, Danny C.  
: APPLICANT: Beck, James J.  
: APPLICANT: Duesing, John H.  
: APPLICANT: Friedrich, Leslie B.  
: APPLICANT: Goodman, Robert M.  
: APPLICANT: Harms, Christian  
: APPLICANT: Means, Jr., Frederick  
: APPLICANT: Montoya, Alice  
: APPLICANT: Meyer, Mary B.  
: APPLICANT: Neuhaus, Jean-Marc  
: APPLICANT: Payne, George B.  
: APPLICANT: Spertson, Christoph

APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Uknes, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Sherricca C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/454,876  
 FILING DATE: 31-MAY-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/181,271  
 FILING DATE: 13-JAN-94  
 APPLICATION NUMBER: US 08/093,301  
 FILING DATE: 16-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/937,197  
 FILING DATE: 6-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/678,378  
 FILING DATE: 1-APR-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/305,566  
 FILING DATE: 6-FEB-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/165,667  
 FILING DATE: 8-MAR-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/042,847  
 FILING DATE: 6-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/632,441  
 FILING DATE: 21-DEC-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/425,504  
 FILING DATE: 20-OCT 1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/848,506  
 FILING DATE: 6-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/768,122  
 FILING DATE: 27-SEP-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/580,431  
 FILING DATE: 7-SEP-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/368,672  
 FILING DATE: 20-JUN-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/045,957  
 FILING DATE: 12-APR-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Elmer, James Scott  
 REGISTRATION NUMBER: 36,119  
 REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1727  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (919)541-8614

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUSER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,876  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/181,271  
FILING DATE: 13-JAN-94  
APPLICATION NUMBER: US 08/093,301  
FILING DATE: 16-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/937,197  
FILING DATE: 6-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/678,378  
FILING DATE: 1-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,139  
REFERENCE/DOCKET NUMBER: S-19825/PL/CGC 1722  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614



RESULT 12  
 US-08-456-262-5/c  
 Sequence 5, Application US/08456262  
 Patent No. 5851766  
 GENERAL INFORMATION:  
 APPLICANT: Ryals, John A.  
 APPLICANT: Alexander, Danny C.  
 APPLICANT: Duesing, John H.  
 APPLICANT: Friedrich, Leslie B.  
 APPLICANT: Goodman, Robert M.  
 APPLICANT: Harms, Christian  
 APPLICANT: Meins, Jr., Frederick  
 APPLICANT: Montoya, Alice  
 APPLICANT: Moyer, Mary B.  
 APPLICANT: Neuhaus, Jean-Marc  
 APPLICANT: Payne, George B.  
 APPLICANT: Sperison, Christoph  
 APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Uknes, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Shericca C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/456,262  
 FILING DATE: 31-MAY-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/181,271  
 FILING DATE: 13-JAN-94  
 APPLICATION NUMBER: US 08/093,301  
 FILING DATE: 16-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/937,197  
 FILING DATE: 6-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/678,378  
 FILING DATE: 1-APR-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/305,566  
 FILING DATE: 6-FEB-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/165,667  
 FILING DATE: 8-MAR-1988  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/042,847  
 FILING DATE: 6-APR-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/632,441  
 FILING DATE: 21-DEC-1990  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/425,504  
 FILING DATE: 20-OCT-1989  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/848,506  
 FILING DATE: 6-MAR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/768,122

```

1      FILING DATE: 27-SEP-1991
2      PRIOR APPLICATION DATA:
3      APPLICATION NUMBER: US 07/580,431
4      FILING DATE: 7-SEP-1990
5      PRIOR APPLICATION DATA:
6      APPLICATION NUMBER: US 07/368,672
7      FILING DATE: 20-JUN-1989
8      PRIOR APPLICATION DATA:
9      APPLICATION NUMBER: US 07/329,018
10     FILING DATE: 24-MAR-1989
11     PRIOR APPLICATION DATA:
12     APPLICATION NUMBER: US 08/045,957
13     FILING DATE: 12-APR-1993
14     ATTORNEY/AGENT INFORMATION:
15     NAME: Elmer, James Scott
16     REGISTRATION NUMBER: 36,129
17     REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1722
18     TELECOMMUNICATION INFORMATION:
19     TELEPHONE: (919)541-8614
20     TELEFAX: (919)541-8689
21     INFORMATION FOR SEO ID NO: 5:
22     SEQUENCE CHARACTERISTICS:
23     LENGTH: 4483 base pairs
24     TYPE: nucleic acid
25     STRANDEDNESS: single
26     TOPOLOGY: linear
27     MOLECULE TYPE: CDNA
28     US-08-456-262-5

```

Query Match	9.2%	Score 33.6;	DB 4;	length 4483;
Best Local Similarity	52.1%;	Pred. NO. 1.1;		
Matches 75;	Conservative 0;	Mismatches 69;	Indels 0;	Gaps 0;

Oy	1	GCTCTTCTGCCCCCTCCCTGAAATTTAATCTTTCACTTCAGTAATTCGACAGATTCACA	60
Db	4481	GCTCTTTTGGCGCTCTGAAATGAGCTCTCTTCTCGCTTTGGAGATGATGAATATGGA	4422
Oy	61	CATTTCAGCTGAGATGATTAATCTGCTTCGCACAAAAAAAAGCTCTTGTTTAAAAATAC	120
Db	4421	CAATTCACAGGAGACTAGTTGTGGGTAAAGATCAAAATAGATCTCTTACCTGGAAGATTAT	4362
Oy	121	TGTGTTGTGAGATCAATCTTGCTT	144
Db	4361	CCCGTTTTGTGTGGCAGCTCTCTT	4338

RESULT 13  
 US-08-456-240-5/c  
 Sequence 5, Application US/08456240  
 Patent No. 5856154  
 GENERAL INFORMATION:  
 APPLICANT: Ryals, John A.  
 APPLICANT: Alexander, Danny C.  
 APPLICANT: Beck, James J.  
 APPLICANT: Duesing, John H.  
 APPLICANT: Friedrich, Leslie B.  
 APPLICANT: Goodman, Robert M.  
 APPLICANT: Harms, Christian  
 APPLICANT: Meins, Jr., Frederick  
 APPLICANT: Montoya, Alice  
 APPLICANT: Moyer, Mary B.  
 APPLICANT: Neuhaus, Jean-Marc  
 APPLICANT: Payne, George B.  
 APPLICANT: Sperison, Christoph  
 APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Ukens, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Shericca C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:

```

ADDRESSSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
City: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,240
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/633,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT 1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4483 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna

```

US-08-456-240-5

Query Match	9.28;	Score 33.6;	DB 4;	length 4483;
Best Local Similarity	52.18;	Pred. NO. 1.1;		
Matches 75;	Conservative 0;	Mismatches 69;	Indels 0;	Gaps 0;

[illegible]

RESULT 14  
US-08-455-736-5/c

- ; Sequence 5, Application US/08455736
- ; Patent No. 5880328

GENERAL INFORMATION:

APPLICANT: Ryals, John A.  
 APPLICANT: Alexander, Danny C.  
 APPLICANT: Beck, James J.  
 APPLICANT: Duesting, John H.  
 APPLICANT: Friedrich, Leslie B.  
 APPLICANT: Goodman, Robert M.  
 APPLICANT: Harms, Christian  
 APPLICANT: Meins, Jr., Frederick  
 APPLICANT: Montoya, Alice  
 APPLICANT: Moyer, Mary B.  
 APPLICANT: Neunhaus, Jean-Marc  
 APPLICANT: Payne, George B.  
 APPLICANT: Sperison, Christoph  
 APPLICANT: Stinson, Jeffrey R.  
 APPLICANT: Uknes, Scott J.  
 APPLICANT: Ward, Eric R.  
 APPLICANT: Williams, Shericca C.  
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC  
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF  
 NUMBER OF SEQUENCES: 106  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: CIBA-GEIGY Corporation  
 STREET: 7 Skyline Drive  
 CITY: Hawthorne  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10532  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/455,736  
 FILING DATE: 31-MAY-1995  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 08/181,271  
 FILING DATE: 13-JAN-1994  
 APPLICATION NUMBER: US 08/093,301  
 FILING DATE: 16-JUL-1993  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/937,197  
 FILING DATE: 6-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/678,378  
 FILING DATE: 1-APR-1991  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/305,566  
FILING DATE: 6-FEB-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/165,667  
FILING DATE: 8-MAR-1988  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/042,847  
FILING DATE: 6-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/632,441  
FILING DATE: 21-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/425,504  
FILING DATE: 20-OCT-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/848,506  
FILING DATE: 6-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/768,122  
FILING DATE: 27-SEP-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/580,431  
FILING DATE: 7-SEP-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/368,672  
FILING DATE: 20-JUN-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/329,018  
FILING DATE: 24-MAR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/045,957  
FILING DATE: 12-APR-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Elmer, James Scott  
REGISTRATION NUMBER: 36,129  
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (919)541-8614  
TELEFAX: (919)541-8689  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-455-736-5

Query Match 9.2%; Score 33.6; DB 4; Length 4483;  
Best Local Similarity 52.1%; Pred. No. 1.1;  
Matches 75; Conservative 0; Mismatches 69; Indels 0; Gaps 0;  
QY 1 GCTCTTCTCCTCCCTCGAATTAATTCCTTCACTGCAATTGCAAGATTACA 60  
DB 4481 GCTCTTCTCCTCCCTCGAATTAATTCCTTCACTGCAATTGCAAGATTACA 4422  
QY 61 CATTTCACTGATGATATATTTGTGCAAAAAAAAAAGTCTTTGTTAAATTC 120  
DB 4421 CAATTCACACGAGTAGTGTGGTAAGATCAAAATAGATCTTAGCTGGAAGTTAT 4362  
QY 121 TTGGTTGTGAATCCATCTTGCTT 144  
DB 4361 CCCGTTGTGTGCAATGCTTCTT 4338

RESULT 15  
US-08-455-073A-1/c  
Sequence 1, Application US/08455073A  
Patent No. 5876949  
GENERAL INFORMATION:  
APPLICANT: Gideon Dreyfuss  
APPLICANT: Mikiko C. Stoml

APPLICANT: Yan Zhang  
TITLE OF INVENTION: Fragile X Related Proteins, Compositions And Methods  
TITLE OF INVENTION: OF Making And Using The Same  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949-15  
STREET: One Liberty Place, 46th floor  
CITY: Philadelphia  
STATE: PA  
COUNTRY: USA  
ZIP: 19103  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Wordperfect 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/455,073A  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Deluca, Mark  
REGISTRATION NUMBER: 33,229  
REFERENCE/DOCKET NUMBER: UPN-2201  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-3100  
TELEFAX: (215) 568-3439  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4362 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 220..2118  
US-08-455-073A-1

Query Match 8.3%; Score 30.4; DB 4; Length 4362;  
Best Local Similarity 55.8%; Pred. No. 7.8;  
Matches 58; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 90 AAAAAAAAAAGTGTCTTTGTTAAATTAATTTACTGTGTTGTAATCCATCTTCTTTC 149  
DB 3322 AAAAAAAAAAGTGTCTTTGTTAAATTAATTTACTGTGTTGTAATCCATCTTCTTTC 3263  
QY 150 CCATTGGAAGTATGATTAACCATCTCTGAAGTGAAGAAA 193  
DB 3262 CAATTCCTGCTTAACATTAAGACATATTTCAAGTTTACAA 3219

Search completed: September 25, 1999, 07:29:06  
Job time: 3383 sec



---

This Page Blank (uspto)

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:02 : Search time 1811.29 Seconds  
(without alignments) 398.382 Million cell updates/sec

Title: US-09-030-606-115

Sequence: 1 GCCTCTTCTCTCCCTCCCTC.....CTGTGACTTGACTTATGTC 366

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	365	99.7	455	AA972883	AA972883 op24g07.s
C 2	365	99.7	538	AI620178	AI620178 tu54a11.x
C 3	363.4	99.3	435	AA809587	AA809587 nj43e11.s
C 4	354	96.7	599	AA813266	AA813266 aj44g03.s
C 5	352	96.2	480	AA437224	AA437224 zv54b10.s
C 6	342	93.4	423	AI685682	AI685682 t89h03.s
C 7	313	85.5	419	AA229495	AA229495 nc35e03.s
C 8	304.4	83.2	411	AA640928	AA640928 nr28d08.r
C 9	299	81.7	383	AI672753	AI672753 we58d10.x
C 10	294.2	80.4	445	AA533772	AA533772 nj93c05.s
C 11	292	79.8	423	AA531606	AA531606 nj65e04.s
C 12	288	78.7	403	AA493522	AA493522 ng73f05.s
C 13	284.2	77.7	339	AA650104	AA650104 ns92f11.s
C 14	281	76.8	377	AI669511	AI669511 wb88f07.x
C 15	280	76.5	374	AA7380	AA7380 zc39h11.s1
C 16	280	76.5	356	AA507804	AA507804 ng89g12.s
C 17	279	76.2	366	AI424280	AI424280 tf01g02.x
C 18	278	76.0	386	AI027196	AI027196 cv83g09.x
C 19	273	74.6	380	AA833543	AA833543 aj36b07.s
C 20	271.8	74.3	382	AI692186	AI692186 wd37g02.x
C 21	270	73.8	371	AA758204	AA758204 ah68e11.s
C 22	266.4	72.8	353	AA508806	AA508806 nb69c06.s
C 23	254.4	69.5	354	AA603372	AA603372 np06g01.s
C 24	248.8	68.0	337	AA400418	AA400418 zu69g10.s
C 25	241.8	66.1	282	AA654919	AA654919 nt77e01.s
C 26	241.4	66.0	342	AA228941	AA228941 nc35h08.s
C 27	225	61.5	472	AA640241	AA640241 nr20h08.s
C 28	223.6	61.1	261	AA650230	AA650230 ns93c02.s
C 29	199.8	54.6	311	AA527743	AA527743 ng95e05.s
C 30	192.4	52.6	346	AI701376	AI701376 wb91f07.x
C 31	172	47.0	283	AA652376	AA652376 ns52h09.s
C 32	159	43.4	267	AA517913	AA517913 nm42f12.s
C 33	102.2	27.9	339	AA917791	AA917791 om82c01.s
C 34	100	27.3	196	AA579219	AA579219 nf36d08.s
C 35	99	27.0	215	AI201254	AI201254 qf70h05.x
C 36	92	25.1	245	AA622798	AA622798 np77b06.s
C 37	90	24.6	157	AI1218326	AI1218326 qh16d03.x
C 38	90	24.6	324	AI423843	AI423843 lg35c04.x
C 39	89	24.3	292	AI695410	AI695410 we79f06.x
C 40	88	24.0	415	AI052805	AI052805 cz11f08.x
C 41	87	23.8	543	AA528104	AA528104 nj16e05.s
C 42	87	23.8	504	AA531255	AA531255 nj09e12.s
C 43	87	23.8	376	AA673355	AA673355 zj58e10.s
C 44	87	23.8	414	AI139031	AI139031 qc22e12.x
C 45	87	23.8	277	AI424769	AI424769 lf09g07.x

#### ALIGNMENTS

RESULT 1  
LOCUS AA972883/c 455 bp mRNA  
DEFINITION OP24g07.s1 Soares\_NFL\_T\_GBC\_S1 Homo sapiens cDNA clone  
ACCESSION AA972883  
IMAGE:1577820 3', mRNA sequence.  
NID G3148063  
VERSION AA972883.1 GI:3148063

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 455)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL  
On Jan 9, 1998 this sequence version replaced gi:937931.  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium ([info@image.lnl.gov](mailto:info@image.lnl.gov)) for further information.  
Insert Length: 633 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 443.  
Location/Qualifiers  
1..455  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:1577820"  
/clone\_lib="Soares\_NFL\_T\_GBC\_S1"  
/lab\_host="DH10B"  
/note="Organ: pooled; Vector: p7T73D-Pac (Pharmacia) with  
a modified polylinker; Site\_: Not I; Site\_2: Eco RI;  
Equal amounts of plasmid DNA from three normalized  
libraries (fetal lung NBHL19w, testis NHT, and B-cell  
NCI-CGAP\_GCB1) were mixed, and ss circles were made in  
vitro. Following HAP purification, this DNA was used as  
tracer in a subtractive hybridization reaction. The driver  
was PCR-amplified cDNAs from pools of 5,000 clones made  
from the same 3 libraries. The pools consisted of  
I.M.A.G.E. clones 297480-302087, 682632-687239,  
726408-728711, and 729096-731399. Subtraction by Bento  
Soares and M. Fatima Bonaldo."  
BASE COUNT 162 a 68 c 94 g 131 t  
ORIGIN  
Query Match 99.7%; Score 365; DB 40; Length 455;  
Best Local Similarity 100.0%; Pred. No. 8.7e-74;  
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 CTCCTTCTCCCTCCTCTGGAATTAATCTTCAACTTGCAATTGCAAGATTACAC 61  
DB 453 CTCCTTCTCCCTCCTCTGGAATTAATCTTCAACTTGCAATTGCAAGATTACAC 394  
QY 62 ATTTCATGATGATATATGTGTGCAAAAAAAGTGTGTTTAAATTAATTAAT 121  
DB 393 ATTTCATGATGATATATGTGTGCAAAAAAAGTGTGTTTAAATTAATTAAT 334  
QY 122 TGGTTTGAATCCATCTGCTTTTCCCATTTGGAATGATCATTAACCATCTCGAA 181  
DB 333 TGGTTTGAATCCATCTGCTTTTCCCATTTGGAATGATCATTAACCATCTCGAA 274  
QY 182 CTGGTAGAAAAACATCTGAAGAGCTAGTCTATCAGATCTGACAGTGAATTGGATGTT 241  
DB 273 CTGGTAGAAAAACATCTGAAGAGCTAGTCTATCAGATCTGACAGTGAATTGGATGTT 214  
QY 242 CTCGAAACATTTACCCAGACGCTGTTCTATCTGTTTAATTAATTAATTAAT 301  
DB 213 CTCGAAACATTTACCCAGACGCTGTTCTATCTGTTTAATTAATTAATTAAT 154  
QY 302 CTCGACATGATACAAACCTGCTCAATCTGCATATAAAGTCTGTGACTGAAGTT 361  
DB 153 CTCGACATGATACAAACCTGCTCAATCTGCATATAAAGTCTGTGACTGAAGTT 94  
QY 362 TACTC 366  
|||||

DB 93 TACTC 89  
RESULT 2  
A1620178/c 538 bp mRNA EST 21-APR-1999  
LOCUS tu54a11.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2254844 3',  
DEFINITION mRNA sequence.  
ACCESSION A1620178  
NID 94629304  
VERSION A1620178.1 GI:4629304  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 538)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
TITLE Tumor Gene Index  
JOURNAL  
On Mar 10, 1998 this sequence version replaced gi:2949075.  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: M. Bento Soares, Ph.D.  
DNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www.bio.lnl.gov/btrp/image/image.html](http://www.bio.lnl.gov/btrp/image/image.html)  
Seq primer: -40UP from Glibco  
High quality sequence stop: 456.  
Location/Qualifiers  
1..538  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="IMAGE:2254844"  
/clone\_lib="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: p7T73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI\_CGAP\_Pr28 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonoids  
985608-986759, 1101192-1101959, and 1217928-1220615)."  
Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 190 a 84 c 113 g 151 t  
ORIGIN  
Query Match 99.7%; Score 365; DB 49; Length 538;  
Best Local Similarity 100.0%; Pred. No. 8.5e-74;  
Matches 365; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 CTCCTTCTCCCTCCTCTGGAATTAATCTTCAACTTGCAATTGCAAGATTACAC 61  
DB 450 CTCCTTCTCCCTCCTCTGGAATTAATCTTCAACTTGCAATTGCAAGATTACAC 391  
QY 62 ATTTCATGATGATATATGTGTGCAAAAAAAGTGTGTTTAAATTAATTAAT 121  
DB 390 ATTTCATGATGATATATGTGTGCAAAAAAAGTGTGTTTAAATTAATTAAT 331  
QY 122 TGGTTTGAATCCATCTGCTTTTCCCATTTGGAATGATCATTAACCATCTCGAA 181  
|||||

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 12:05:07 ; Search time 1811.29 seconds  
(without alignments)  
1377.612 Million cell updates/sec

Title: US-09-030-606-173  
Perfect score: 1265  
Sequence: 1 GGCAGCCCGCAGCTCGCAGCC.....AGAGANGNCAGCAAAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*  
42: em\_est42:\*  
43: em\_est43:\*  
44: em\_est44:\*  
45: em\_est45:\*  
46: em\_est46:\*  
47: em\_est47:\*  
48: em\_est48:\*  
49: em\_est49:\*  
50: em\_est50:\*  
51: em\_est51:\*  
52: em\_est52:\*  
53: em\_est53:\*

54: em\_est22:\*  
55: em\_est23:\*  
56: em\_est24:\*  
57: em\_est25:\*  
58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	476.8	37.7	777	48	AI557281	PT2.1_15
2	355.2	28.1	404	34	AA503963	AA503963 n39601.s
3	315.4	24.9	559	50	AI686689	tU5911.x
4	305.6	24.2	415	35	AA551449	AA551449 n155605.s
5	290.2	22.9	576	34	AA533140	AA533140 n14606.s
6	241.2	19.1	722	48	AI557025	PT2.1_10
7	209.2	16.5	259	35	AA565161	AK52C07.s
8	199.8	15.8	229	35	AA552779	AA552779 nK57A10.s
9	192.8	15.2	241	36	AA603977	AA603977 n047A09.s
10	192	15.2	218	50	AI674521	AI674521 wC39602.x
11	176	13.9	216	32	AA336074	EST4086
12	167.2	13.2	188	50	AI675815	wB97B12.x
13	155	12.3	191	35	AA595489	AA595489 n03404.s
14	115.8	9.2	507	35	AA583052	AA583052 n080909.s
15	109.8	8.7	517	31	AA293027	zE54A12.r
16	108.4	8.6	498	26	W73140	W73140 zd55e11.r1
17	108.2	8.6	505	45	AI391329	mb71n07.y
18	101.2	8.0	590	33	AA411252	zE33B03.r
19	101	8.0	321	31	AA302930	EST113097
20	99.6	7.9	333	49	AI636241	tE93C03.x
21	98.8	7.8	686	46	AI415008	mb71n07.x
22	97.4	7.7	586	26	W73140	W73140 zd55e11.s1
23	96	7.6	356	39	AA838788	a187G08.s
24	95.6	7.6	504	49	AI620091	LY48A10.x
25	93.2	7.4	212	28	C16642	C16642 C16642 Clon
26	91.4	7.2	420	43	AI226226	ue88F06.y
27	90.8	7.2	367	41	AI002101	CL38403.s
28	90	7.1	400	31	AA293231	zE26609.r
29	90	7.1	539	34	AA477689	zU44A12.r
30	89.8	7.1	480	39	AA846771	aJ41F01.s
31	89	7.0	234	33	AA452459	zX23909.r
32	86.8	6.9	519	33	AA401397	zU68B01.s
33	86	6.8	122	49	AI632842	tE32H08.x
34	85.4	6.8	396	49	AV005299	AV005299 AV005299
35	85.2	6.7	493	45	AI324874	m185B05.x
36	85.2	6.7	363	47	AI477594	IB57F08.x
37	84.4	6.7	392	25	W16362	W16362 mb56D06.r1
38	84.4	6.7	457	45	AI385433	m185B05.y
39	84.2	6.7	1029	25	W13212	me83A07.r1
40	84	6.6	499	43	AI178581	EST23260
41	83.8	6.6	626	36	AA643312	n559B08.s
42	83.6	6.6	436	42	AI139437	QC20E03.x
43	83.2	6.6	585	35	C23111	C23111 Japa
44	82.8	6.5	599	43	AI237604	EST234166
45	82.8	6.5	440	46	AI411685	EST239979

#### ALIGNMENTS

RESULT 1  
LOCUS AI557281 777 bp mRNA  
DEFINITION PT2.1\_15.G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.  
ACCESSION AI557281  
NID 94489644  
VERSION AI557281.1 GI:4489644  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 777)  
AUTHORS Huang, G.M., Ng, W., Farakas, J., Chen, L., Liang, H.A., Gordon, D., Jun Yu, J., and Hood, L.  
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis  
JOURNAL Unpublished (1999)  
COMMENT On May 18, 1998 this sequence version replaced g1:3138767.

Contact: Guyang Mathew Huang  
Leroy Hood  
University of Washington  
Department of Molecular Biotechnology, Box 357730, University of  
Washington, Seattle, WA 98195  
Tel: 5106280100  
Fax: 5106280108  
Email: huangm@yahoo.com.

FEATURES  
source  
1..777  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="960H11; 6: 6p21.31-6p22.1; 21q"  
/clone\_lib="tumor2"  
/note="Organ: Prostate; Vector: pBluescript; Directional  
cDNA library was constructed using Lambda ZP II kit  
(Stratagene). mRNA was extracted from a frozen prostate  
tumor tissue (Mayo Clinics)."

BASE COUNT 162 a 205 c 227 g 157 t 26 others  
ORIGIN

Query Match 37.7%; Score 476.8; DB 48; Length 777;  
Best Local Similarity 81.9%; Pred. No. 9.6e-106;  
Matches 643; Conservative 0; Mismatches 48; Indels 94; Gaps 5;

8 CGCAGCTGCGACCCCTGGCAGGCGGACATGGTCATGAGAAAGATTTCTGCTGGGCG 67  
15 CGCAGCTGCGACCCCTGGCAGGCGGACATGGTCATGAGAAAGATTTCTGCTGGGCG 74  
68 TCTGTGTCATCCGACAGTGGGTGCTGTCAAGCGGACACATGTTTCCAGAACTCTACACCA 127  
75 TCTGTGTCATCCGACAGTGGGTGCTGTCAAGCGGACACATGTTTCCAGAACTCTACACCA 134  
128 TCGGGCTGGGCTGACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATGGTGAGG 187  
135 TCGGGCTGGGCTGACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATGGTGAGG 194  
188 CGAGCTCTCCGTCAGGCGGACAGTACAGACCTTGTGCTAAGAGACCTCATGC 247  
195 CGAGCTCTCCGTCAGGCGGACAGTACAGACCTTGTGCTAAGAGACCTCATGC 254  
248 TCATCAAGTTGAGAGATCCGTGTCCAGTCTGACACCATCCGAGCATCAGATTCTT 307  
255 TCATCAAGTTGAGAGATCCGTGTCCAGTCTGACACCATCCGAGCATCAGATTCTT 314  
308 CGCAGTCCCTACCGCGGGGAATCTTGCTTGTGGCTGGGGTCTGCTGGCAAG 367  
315 CGCAGTCCCTACCGCGGGGAATCTTGCTTGTGGCTGGGGTCTGCTGGCAAG 374  
368 GTGAGCTCAGGGGTGTGTGTGCTGCTTCAAGAGGTCTCTGCCAAGTCCGGGGGGC 427  
375 G----- 376

428 TGACCCAGAGCTTGCCTCCAGGAGAGATGCTTACCGTGTGAGTGCATGTC 487  
376 -----CAGAAATGCTACGCTGTGTCAGTGTGAGTGCATGTC 410  
488 GGTGGTCTGAGAGAGTGTGAGTAACTCTATGACCCGCTGACCAACCCAGATGT 547  
411 GGTGGTCTGAGAGAGTGTGAGTAACTCTATGACCCGCTGACCAACCCAGATGT 470

QY 548 CTGCGCGGCGGAGGCGCAAGAGAGACTCTGCAAGGAGTCTGGGGGGCCCT 607  
DB 471 CTCGCGCGGCGGAGGCGCAAGAGAGACTCTGCAAGGAGTCTGGGGGGCCCT 529  
QY 608 GATCTCAACGGGTACTTGCAGGCGCTTGTCTTTCGAAAGCCCGTGTGGCCAGT 667  
DB 530 GATGTGAAACGGGCGGCTTGCAGGCGCTTGTCTTTCGAAAGCCCGTGTGGCCAGT 587  
QY 668 TGGCTGCCAGGTGTGTACCAACCTCTGCAATTCATGAGTGTGAGAAAACCT 727  
DB 588 TGGCTGCCAGGTGTGTACCAACCTCTGCAATTCATGAGTGTGAGAAAACCT 642  
QY 728 CCAGCCAGTACTTGGGAGTGGGAGCAACCATGAAATGACCCCAATCATCTGC 787  
DB 643 C--AGGACAGTTACTTGGGAGTGGGAGCAACCATGAAATGACCCCAATCATCTGC 700  
QY 788 GGAAG 792  
DB 701 GAANG 705

RESULT 2  
AA503963/C  
LOCUS  
DEFINITION  
AA503963 404 bp mRNA  
nh39a01.s1 NCI-CGAP\_P15 Homo sapiens cDNA clone IMAGE:954696  
similar to SW:KLR\_MOUSE P15946 GLANDULAR KALLIKREIN K11 PRECURSOR  
; contains MSRL.b2 MSRL repetitive element; mRNA sequence.  
AA503963  
92238930  
VERSION  
AA503963.1 GI:2238930  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 404)  
AUTHORS  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
On May 8, 1995 this sequence version replaced g1:801235.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.L.Strausberg@nih.gov  
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuequi,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 1244 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 144.

FEATURES  
source  
1..404  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:954696"  
/clone\_lib="NCI-CGAP\_P15"  
/sex="male"  
/tissue\_type="prostate"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; mRNA made from normal prostatic  
epithelial cells, cDNA made by oligo-dT priming.  
Non-directionally cloned. Size selected on agarose gel,  
average insert size 600 bp.

BASE COUNT 64 a 91 c 159 g 90 t



REFERENCE 1 (bases 1 to 415)  
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 JOURNAL Tumor Gene Index  
 COMMENT Unpublished (1997)  
 On May 9, 1995 this sequence version replaced gi:802243.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,  
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center.  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbdp/image/image.html](http://www.bio.llnl.gov/dbdp/image/image.html)

Insert Length: 640 Std Error: 0.00  
 Seq primer: -40m3 fwd. ET from Amersham  
 High quality sequence stop: 412.

## FEATURES

source Location/Qualifiers

1..415  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:996416"  
 /clone\_1lb="NCI-CGAP\_Prg9"  
 /sex="male"  
 /tissue\_type="normal prostatic epithelial cells"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: PAMP10; mRNA made from  
 normal prostatic epithelial cells. cDNA made by oligo-dT  
 priming. Non-directionally cloned. Size-selected on  
 agarose gel, average insert size 600 bp. Library made by  
 D. Kitzman, NIH."

BASE COUNT 78 a 125 c 129 g 83 t

## ORIGIN

Query Match 24.2%; Score 305.6; DB 35; Length 415;  
 Best Local Similarity 81.3%; Pred. No. 2.2e-64;  
 Matches 405; Conservative 0; Mismatches 9; Indels 84; Gaps 1;

```

QY 149 TTGAGCGCGCAGAGCGGAGCCAGATGTGGAGCGGCACCTCTCGTAGCGGACC 208
DB 2 TCACGCCGCGCAGAGCGGAGCCAGATGTGGAGCGGCACCTCTCGTAGCGGACC 61
QY 209 CAGAGTAGCAACAGACCCCTGCTGCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCG 268
DB 62 CAGAGTAGCAACAGACCCCTGCTGCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCG 121
QY 269 TGTCCAGATGTGACACCATCCGAGCATCAGCATTTGCTTGGCAGTCCCTAACCGCGGGA 328
DB 122 TGTCCAGATGTGACACCATCCGAGCATCAGCATTTGCTTGGCAGTCCCTAACCGCGGGA 181
QY 329 ACTCTTGCCCTGTTTGGCTGGGGTCTGCTGGGGAAGGCTGAGCTACGGGTGTGTCTC 388
DB 182 ACTCTTGCCCTGTTTGGCTGGGGTCTGCTGGGGAAGGCTGAGCTACGGGTGTGTCTC 222
QY 389 TGCCCTTTTCAAGAGAGTCTCTGCCAGTCGGGGGGCTGACCACAGACTCTGCGTCCC 448
DB 222 ----- 222
QY 449 AGGCAGAAAGCTACCGTGTCTGTCAGTGTGTAACGTCGTGGGTGTCTGAGAGAGTCTG 508
DB 222 -----CAGAAATGCTACCGTGTCTGTCAGTGTGTAACGTCGTGGGTGTCTGAGAGAGTCTG 277
QY 509 CAGTAAGCTTATGAGCCGCTGTACACACCCAGCATGTTTGGCGCGCGGAGGCGCAAGA 568
DB 278 CAGTAAGCTTATGAGCCGCTGTACACACCCAGCATGTTTGGCGCGCGGAGGCGCAAGA 337
QY 569 CCAGAAAGAGACTCTGCAACGAGTGTGGGGGGCCCTGATCTGCAACGGGTACTTGCA 628

```

```

DB 338 CCGAAGAGACTCTCTGCAACGAGTGTGGGGGGCCCTGATCTGCAACGGGTACTTGCA 397
QY 629 GGGCCTTGTGTCTTTCCG 646
DB 398 GGGCCTTGTGTCTTTCCG 415

```

## RESULT 5

AA533140

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,  
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 DNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbdp/image/image.html](http://www.bio.llnl.gov/dbdp/image/image.html)

Insert Length: 885 Std Error: 0.00  
 Seq primer: -40m3 fwd. ET from Amersham  
 High quality sequence stop: 256.

## FEATURES

source Location/Qualifiers

1..576  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="IMAGE:995579"  
 /clone\_1lb="NCI-CGAP\_Prg9"  
 /sex="male"  
 /tissue\_type="normal prostatic epithelial cells"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: PAMP10; mRNA made from  
 normal prostatic epithelial cells. cDNA made by oligo-dT  
 priming. Non-directionally cloned. Size-selected on  
 agarose gel, average insert size 600 bp. Library made by  
 D. Kitzman, NIH."

BASE COUNT 169 a 156 c 139 g 102 t 10 others

## ORIGIN

Query Match 22.9%; Score 290.2; DB 34; Length 576;  
 Best Local Similarity 95.7%; Pred. No. 1.3e-60;  
 Matches 309; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

```

QY 567 GACCAAGAGACTCTGCAACGAGTACT-CTGGGGGGCCCTGATCTGCAACGGGTACTT 625
DB 42 GACCAAGAGACTCTGCAACGAGTACTCTGTTGGGGCCCTGATCTGCAACGGGTACTT 101
QY 626 GCAGGGCCTTGTGTCTTTTGGAAAGCCCGCTGTGGCCAAAGTTGGCGTCCAGAGTCTTA 685
DB 102 GCAGGGCCTTGTGTCTTTTGGAAAGCCCGCTGTGGCCAAAGTTGGCGTCCAGAGTCTTA 161

```



```

OY 686 CACCAACCTCTGCAAAATTCAGTAGTGATAGAGAAAAACCGTCAGCCAGTTAACTCTG 745
      |||||||
Db 162 CACCAACCTCTGCAAAATTCAGTAGTGATAGAGAAAAACCGTCAGCCAGTTAACTCTG 221
      |||||||
OY 746 GGCACCTGGGAACCAATTAATGACCCCAATATCATCTCTGGGAAAGATTGAGAAATA 805
      |||||||
Db 222 GGCACCTGGGAACCAATTAATGACCCCAATATCATCTCTGGGAAAGATTGAGAAATA 281
      |||||||
OY 806 TCTGTTCCCAAGCCCTCTCTCTCAGGCGCCAGAGATCCAGGCCGCCAGCCCTCTCTCT 865
      |||||||
Db 282 TCTGTTCCCAAGCCCTCTCTCTCAGGCGCCAGAGATCCAGATCCCAAGAACTACTTCT 341
      |||||||
OY 866 CAACCAAGGATACAGATCCCA 888
      |||||
Db 342 TCAACCAAGGATACAGATCCCA 364
      |||||

RESULT 6
LOCUS A1557025 722 bp mRNA EST 23-MAR-1999
DEFINITION PT2.1_10_F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.
ACCESSION A1557025
NID 94489388
VERSION A1557025.1 GI:4489388
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 722)
AUTHORS Huang,G.M., Ng,W., Farakas,J., Chen,L., Liang,H.A., Gordon,D., Jun
Yu,J. and Hood,L.
TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

Contact: Guyang Mathew Huang
Leroy Hood
University of Washington
Department of Molecular Biotechnology, Box 357730, University of
Washington, Seattle, WA 98195
Tel: 5106280100
Fax: 5106280108
Email: huangm@yahoo.com

FEATURES
source
1..722
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="15: 21q"
/clone_11b="tumor2"
/note="Organ: Prostate; Vector: pBluescript; Directional
cDNA library was constructed using Lambda ZP II kit
(Stratagene). mRNA was extracted from a frozen prostate
tumor tissue (Mayo Clinics)."
```

BASE COUNT 146 a 156 c 195 g 176 t 49 others

Query Match 19.1%; Score 241.2; DB 48; Length 722;  
Best Local Similarity 89.7%; Pred. No. 1.1e-48;  
Matches 297; Conservative 0; Mismatches 26; Indels 8; Gaps 4;

```

OY 115 AACTCTACACATGGGCTGGGCTGACAGTCTTAGAGCCGACCAAGACCGAGGAGC 174
      |||||||
Db 115 AGCTTCTACACATGGGCTGGGCTGACAGTCTTAGAGCCGACCAAGACCGAGGAGC 174
      |||||||
OY 175 CAGATGGTGGAGGCACCTCTCGTAGCGGACAGACAGACAGACACCTTGTGGCT 234
      |||||||
Db 175 CAGATGGTGGAGGCACCTCTCGTAGCGGACAGACAGACAGACACCTTGTGGCT 234
      |||||||
OY 235 AACGACCTCATGCTCAATGATGAGAAATCCGTGTCCAGTCTGACACCATCGGAGC 294
      |||||||
```

```

Db 235 AACGACCTCATGCTCAATGATGAGAAATCCGTGTCCAGTCTGAAACCATCGGAGC 294
      |||||||
OY 295 ATGAGATTGCTTCGCGAGTGGCCCTACCGCGGGGAACCTTGGC--TGTGTTGGCTGGG 353
      |||||||
Db 295 ATGAGATTGCTTCGCGAGTGGCCCTACCGCGGGGAACCTTTCCTGTTGGCTGGG 354
      |||||||
OY 354 TCTGTC--TGGCGAAGGTGAGCTCAGCGGTGTGTGTCTGCTCTTC----AAGAGGTCC 408
      |||||||
Db 355 TCTGTTGGCGGAACGGTGAAGCTCAGCGGTGTGTGTCTGCTCTTC--TAAAGAGGTCC 414
      |||||||
OY 409 TCTGCCCA--GTGCGCGGGGCTACCCACAG 437
      |||||
Db 415 TTTTNCNAAGTNTGCGGGGCTNACCCAGAG 445
      |||||
```

```

RESULT 7
LOCUS AA565161/c 259 bp mRNA EST 08-SEP-1997
DEFINITION nk52c07.s1 NCI_CGAP_P7 Homo sapiens cDNA clone IMAGE:1017132
similar to contans MSRL.b3 MSRL MSRL repetitive element ;, mRNA
sequence.
ACCESSION AA565161
NID 92336800
VERSION AA565161.1 GI:2336800
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 259)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1397569.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagui,
M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/URL at:
www-bio.llnl.gov/dbp/image/image.html
```

Insert Length: 458 Std Error: 0.00  
Seq primer: -40m3 fwd. ET from Amersham  
High quality sequence stop: 258.

```

FEATURES
source
1..259
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="17q21"
/clone_11b="IMAGE:1017132"
/clone_11b="NCI_CGAP_P7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH10B"
/note="Organ: prostate; Vector: PAMP10; mRNA made from
prostate intraepithelial neoplasia (high-grade), cDNA
made by oligo-dt priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp."

BASE COUNT 56 a 41 c 106 g 56 t
```

Query Match 16.5%; Score 209.2; DB 35; Length 259;  
Best Local Similarity 88.8%; Pred. No. 4.2e-41;  
Matches 231; Conservative 0; Mismatches 28; Indels 1; Gaps 1;

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuagut,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbirp/image/image.html](http://www-bio.llnl.gov/dbirp/image/image.html)

Insert Length: 419    Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 215.

Location/Qualifiers  
1. .229

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="1017594"
/clone_id="NCI_CGAP_Pc7"
/sex="male"
/tissue_type="low-grade prostatic neoplasia"
/lab_host="DH10B"
/notes="Organ: prostate; Vector: PAMPl0; mRNA made from prostate intraepithelial neoplasia (high-grade), cDNA made by oligo-dt priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp."

```

49 a 38 c 94 g 48 t

Query Match	15.8%	Score 199.8;	DB 35;	Length 229;
Best Local Similarity	90.4%;	Pred. No. 7.6e-39;		
Matches 207; Conservative	0;	Mismatches 22;	Indels 0;	Gaps 0

QY	1026	CCCCAACCCCCCTCCCTCCCTTCAGATGTCCAGAGTCCAAAGCCCCCAACCCCTGTGTCCCCAGAC	108
Db	229	CCCCAGCCCCCTCCTCCTCCCTCAGAACCCAGAGATCCAAAGCCCCCAACCCCTCCTTCCCCAGAC	170
QY	1086	CCAGAGGTNNAGGTCCAGCCGCCCTCTTCCTCCTAGACCCAGNNGTCCAAATGCCCCCTAAAT	1143
Db	169	CCAGAGTCCAGGTACAGAGCCCTCCTCCCTCCTAGAACCCAGCGGTCCAAATGCCACTAGAC	110
QY	1146	TTTCCCTCGNACACAGTGTCCCTCTTGTGTGNAAGTTGACCCCAACTTACCAAGTTGGTTTTTC	1205
Db	109	TCTCCCTGTACACAGTGTCCCTCTTGTGTGACGTGTGACCCCAACTTACCAAGTTGGTTTTTC	50
QY	1206	ATTTTTNGCCTTTCCCTCCCTAGATCCAGAAATPAAAGTTAAAGANGNG	1254
Db	49	ATTTTGTCTCCCTTCCCTCCCTAGATCCAGAAATPAAAGTTAAAGAAAGCG	1

RESULT	9.
AA603977/c	
LOCUS	AA603977
DEFINITION	nc047809.s1 NC1.CGAP_Pt23 Homo sapiens cDNA clone IMAGE:1103800 3'
	241 bp mRNA EST 28-OCT-1997

ACCESSION	AA603977	
NID	92444547	
VERSION	AA603977.1	GI:2444547
KEYWORDS	EST.	
SOURCE	human.	

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 241)  
 NCBI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 National Cancer Institute, Cancer Genome Anatomy Project (CCAP)  
 Tumor Gene Index  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced g4:1395285.

On Sep 12, 1996 this sequence replaced g1:1395285

Contact: Robert Strausberg, Ph.D  
Tel: (301) 496-1550

Email: Robert\_Strausberg@nh.gov  
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.

CDNA Library Preparation: Stratagene, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.lnl.gov/dbfp/image/image.html](http://www-bio.lnl.gov/dbfp/image/image.html)

Insert Length: 2084    Std Error: 0.00  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 231.

FEATURES

SOURCE

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1103800"
/clone_lib="NCI CGAP_Pcr3"
/sex="male, pooled"
/tissue_type="prostate tumor"
/lab_host="SO4 (kanamycin resistant)"
/notes="Organ: prostate; Vector: Bluescript SK-, Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
0110 dt. Pooled prostate tumors. 5' adaptor sequence: 5'
TAATTCGGCAGACG 3' 3' adaptor sequence: 5'

```

CTCAGATTGTTTTTTTTTTTTTTT 3' Average insert size: 1.2 kb."

BASE COUNT 46 a 37 c 78 g 80 t

ORIGIN

Query Match 15.2% Score 192.8; DB 36; Length 241;  
Best Local Similarity 92.1% Pred. No. 3.8e-37;  
Matches 197; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

1052 GAGGTCCAGGCCCCCAACCCCTGTCCTCCAGACGAGGTNNAGTCCAGCCCTCT 1111  
|||||

Db 241 GAGGTCCAGGCCCCCAACCCCTGTCCTCCAGACGAGGTCCAGCCCTCTCC 182  
|||||

Qy 1112 TCCNTCAGACCCAGNGTCCATGCCACTGATTTTCCTGNAACAGATGCCCTTGT 1171  
|||||

Db 181 TCCCTCAGACCCAGGCTCCATGCCACTGATTTTCCTGNAACAGATGCCCTTGT 122  
|||||

Qy 1172 GGNAGTTGACCAACCTTACCAAGTTGGTTTTCATTTTTCCTTCCCTAGATTC 1231  
|||||

Db 121 GGCACTGTGACCAACCTTACCAAGTTGGTTTTCATTTTTCCTTCCCTAGATTC 62  
|||||

Qy 1232 AGAATAAGTTAAGANGNGCAAAAAA 1265  
|||||

Db 61 AGAATAAGTTAAGANGNGCAAAAAA 28  
|||||

RESULT 10  
A1674521/c 218 bp mRNA EST 19-MAY-1999  
LOCUS A1674521  
DEFINITION WC39902.x1 NCI\_CGAP\_Pr28 Homo sapiens CDNA IMAGE:2321042 3'  
sequence.  
A1674521  
VERSION 94875001  
KEYWORDS A1674521.1 GI:4875001  
EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 218)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Mar 16, 1998 this sequence version replaced gi:2961915.  
COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmett-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www-bio.llnl.gov/dbp/image/image.html

Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 218  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:2321042"  
/clone\_id="NCI\_CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI\_CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,

this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonids  
985608-986759, 1101192-1101959, and 1217928-1220615).  
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 48 a 36 c 85 g 49 t

ORIGIN

Query Match 15.2% Score 192; DB 50; Length 218;  
Best Local Similarity 90.8% Pred. No. 5.8e-37;  
Matches 198; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

1038 CTCCTCAGAGTCCAGAGTCCAGACCCCAACCCCTGTCCTCCAGACCCAGAGTNNAG 1097  
|||||

Db 218 CTCCTCAGAGTCCAGAGTCCAGACCCCAACCCCTGTCCTCCAGACCCAGAGTCCAG 159  
|||||

Qy 1098 GTCCAGCCCTCTTCCTCAGACCCAGACCCCAACCCCTGTCCTCCAGACCCAGAGT 1157  
|||||

Db 158 GTCCAGCCCTCTTCCTCAGACCCAGACCCCAACCCCTGTCCTCCAGACCCAGAGT 99  
|||||

Qy 1158 CAGTGCCTCTTGTGGANAGTCCAGACCCCAACCCCTGTCCTCCAGACCCAGAGT 1217  
|||||

Db 98 CAGTGCCTCTTGTGGANAGTCCAGACCCCAACCCCTGTCCTCCAGACCCAGAGT 39  
|||||

Qy 1218 TTTCCTCAGATCCAGATTAAGTTAAGANGNGC 1255  
|||||

Db 38 TTTCCTCAGATCCAGATTAAGTTAAGANGNGC 1

RESULT 11  
AA336074 216 bp mRNA EST 21-APR-1997  
LOCUS AA336074  
DEFINITION EST40886 Endometrial tumor Homo sapiens CDNA 5' end similar to  
similar to kallikrein family, mRNA sequence.  
AA336074  
VERSION 91988560  
KEYWORDS AA336074.1 GI:1988560  
EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 216)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,  
Bull, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, Wai, C.,  
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghagen, N.S.,  
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,  
Kelly, D.W., Kelley, J.C., Liu, L.-I., Marmaro, S.M., Merrick, D.M.,  
Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Uterback, T.R., Weidman, J.F., Li, Y.,  
Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,  
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Weissner, P.S., Olsen, H.,  
Raymond, L., Wei, Y.F., Wang, J., Xu, C., Yu, G.L., Ruben, S.M.,  
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
Fraser, C.M. and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl.), 3-174 (1995)  
JOURNAL MEDLINE  
COMMENT On Apr 14, 1993 this sequence version replaced gi:592773.  
Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423

Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M3 Reverse

# FEATURES

source  
 Location/Qualifiers  
 1..216  
 /organism="Homo sapiens"  
 /db\_xref="ATCC (inhost):138014"  
 /db\_xref="taxon:9606"  
 /clone\_11b="Endometrial tumor"  
 /sex="female"  
 /dev\_stage="adult"  
 /note="Organ: endometrium; Vector: pBluescript SK-;  
 Site\_1: EcoRI; Site\_2: XhoI"  
 45 a 65 c 63 g 38 t 5 others

## BASE COUNT

ORIGIN

## Query Match

Best Local Similarity 13.9%; Score 176; DB 32; Length 216;  
 Matches 176; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GCAGCCCGACCTGCGACCCCTGCGAGCGGCGACTGTCATGAAAAAGATTGCTGCT 61  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 36 GCAGCCCGACCTGCGAGCCCTGCGAGCGGCGACTGTCATGAAAAAGATTGCTGCT 95  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 62 CGGCGCTCCGTGTCAGTCGCGAGTGGTGTCTCAGCCGCGACCTGTTCCAGACTCC 121  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 96 CGGCGCTCCGTGTCAGTCGCGAGTGGTGTCTCAGCCGCGACCTGTTCCAGACTCC 155  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 122 ACACCATCGGGCTGGGCTGCGACAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATG 181  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 156 ACACCATCGGGCTGGGCTGCGACAGTCTTGAGGCGCAGCAAGAGCCAGGAGCCAGATG 215  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 182 T 182  
 ||  
 DB 216 T 216

RESULT 12  
 LOCUS A1675815/c 188 bp mRNA EST 19-MAY-1999  
 DEFINITION wb97b12.x1 NCI\_CGAP\_P128 Homo sapiens cDNA clone IMAGE:2313599 3',  
 mRNA sequence.  
 ACCESSION A1675815  
 NID 94876295  
 VERSION A1675815.1 GI:4876295  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 188)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Dec 20, 1995 this sequence version replaced gi:1133540.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/birp/image/image.html](http://www.bio.llnl.gov/birp/image/image.html)  
 Seg. primer: 40UP from G1bco.

# FEATURES

source  
 Location/Qualifiers  
 1..188  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_11b="NCI\_CGAP\_P128"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI\_CGAP\_P128 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clones  
 985608-986759, 1101192-1101959, and 1217928-1220615).  
 Subtraction by Bento Soares and M. Fatima Bonaldo."  
 43 a 31 c 61 g 53 t

## BASE COUNT

ORIGIN

Query Match 13.2%; Score 167.2; DB 50; Length 188;  
 Best Local Similarity 92.9%; Pred. No. 5.6e-31;  
 Matches 169; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1084 ACCGAGGTNNAGTCCAGCCCTCTCCCTCAGACCCAGNGGTCGAATGCCACTAG 1143  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 188 ACCGAGGTCCAGGTCCAGCCCTCTCCCTCAGACCCAGGCTCCAAATGCCACTAG 129  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1144 ATTTCCCTGACAGAGCCCTTGTGNNAGTTGACCCACTTACAGTTGTTT 1203  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 128 ACTGCCCTGTACAGAGGCCCTTGTGACAGTTGACCCACTTACAGTTGTTT 69  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1204 TCATTTTNGTCCCTTCCCTAGATCCAGAAATTAAGTTAAGANGGCCAAAAA 1263  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 DB 68 TCATTTTNGTCCCTTCCCTAGATCCAGAAATTAAGTTAAGANGGCCAAAAA 9  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 QY 1264 AA 1265  
 ||  
 DB 8 AA 7

RESULT 13  
 LOCUS AA595489/c 191 bp mRNA EST 18-SEP-1997  
 DEFINITION n034h04.s1 NCI\_CGAP\_P123 Homo sapiens cDNA clone IMAGE:1102615 3',  
 mRNA sequence.  
 ACCESSION AA595489  
 NID 92410839  
 VERSION AA595489.1 GI:2410839  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 1 (bases 1 to 191)  
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Apr 14, 1993 this sequence version replaced gi:692625.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Stratagene, Inc.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:24:11 ; Search time 2928.69 Seconds  
(without alignments)  
1584.350 Million cell updates/sec

Title: US-09-030-606-174

Sequence: 1 GGTCAAGCCGACACGTGTTTC.....TCAAAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl.\*  
1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_pat.\*  
6: gb\_ph.\*  
7: gb\_pl1.\*  
8: gb\_pl2.\*  
9: gb\_pl3.\*  
10: gb\_pr2.\*  
11: gb\_pr3.\*  
12: gb\_ro.\*  
13: gb\_st.\*  
14: gb\_sts.\*  
15: gb\_sy.\*  
16: gb\_un.\*  
17: gb\_v1.\*  
18: em\_fun.\*  
19: em\_hlg.\*  
20: em\_hum1.\*  
21: em\_hum2.\*  
22: em\_in.\*  
23: em\_om.\*  
24: em\_or.\*  
25: em\_ov.\*  
26: em\_pat.\*  
27: em\_ph.\*  
28: em\_pl.\*  
29: em\_ro.\*  
30: em\_sts.\*  
31: em\_sy.\*  
32: em\_un.\*  
33: em\_v1.\*  
34: gb\_hlg1.\*  
35: gb\_hlg2.\*  
36: gb\_in1.\*  
37: gb\_in2.\*  
38: em\_ba1.\*  
39: em\_ba2.\*  
40: em\_hum3.\*  
41: em\_hum4.\*  
42: gb\_pr4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1411.4	96.7	4740	42	AF135023	Homo sapi

2	1411.4	96.7	4385	42	AF148532	AF148532 Homo sapi
3	1405	96.3	5900	42	AF113141	AF113141 Homo sapi
4	296.4	20.3	1347	42	AF113140	AF113140 Homo sapi
5	248.2	17.0	166723	34	HS310013	AL031658 Homo sapi
6	240.6	16.5	151770	10	HSAC002066	AC002066 Homo sapi
7	240	16.4	166663	35	AC007728	AC007728 Homo sapi
8	238.8	16.4	234053	11	AC002429	AC002429 Homo sapi
9	238.6	16.4	142178	35	AC007160	AC007160 Homo sapi
10	236	16.2	179757	42	AC006960	AC006960 Homo sapi
11	235	16.1	134578	34	AC004824	AC004824 Homo sapi
12	235	16.1	220000	35	AC005674	AC005674 Homo sapi
13	234.6	16.1	102200	9	HS30M3	AL031775 Human DNA
14	234.6	16.1	80507	11	AC004024	AC004024 Homo sapi
15	234.6	16.1	188477	34	AC004971	AC004971 Homo sapi
16	234	16.0	240000	35	AC006593	AC006593 Homo sapi
17	233.6	16.0	96625	10	HSAC000118	AC000118 Homo sapi
18	233.6	16.0	135448	34	AC004900	AC004900 Homo sapi
19	232.8	16.0	39760	11	AC005253	AC005253 Homo sapi
20	232.8	16.0	46213	11	AC005387	AC005387 Homo sapi
21	232.8	16.0	42030	42	AC007565	AC007565 Homo sapi
22	232.2	15.9	45997	11	AC000028	AC000028 Homo sapi
23	232.2	15.9	45976	11	HS081031	U81031 Homo sapi
24	232.2	15.9	169026	35	AC007371	AC007371 Homo sapi
25	231.8	15.9	234596	34	HS94M16	AL078600 Homo sapi
26	231.8	15.9	124956	34	HS193N13	AP000046 Homo sapi
27	231.4	15.9	100000	9	AP000046	AP000046 Homo sapi
28	231.4	15.9	100000	9	AP000114	AP000114 Homo sapi
29	231.4	15.9	100000	9	AP000130	AP000130 Homo sapi
30	231.4	15.9	47559	9	AP000303	AP000303 Homo sapi
31	231.4	15.9	43012	35	AC007850	AC007850 Homo sapi
32	231.2	15.8	134970	11	AC002504	AC002504 Homo sapi
33	231.2	15.8	127587	34	AC005959	AC005959 Homo sapi
34	230.8	15.8	208643	34	HS44M10	Z97197 Homo sapi
35	230.6	15.8	82359	42	AC004922	AC004922 Homo sapi
36	230.6	15.8	82359	42	AC004922	AC004922 Homo sapi
37	230.4	15.8	79376	9	HS454G6	Z98750 Homo sapi
38	230.2	15.8	114784	10	HS1189B24	AL030996 Homo sapi
39	230.2	15.8	164028	42	AC004890	AC004890 Homo sapi
40	230	15.8	84170	9	HS330M4	Z75887 Homo sapi
41	230	15.8	20733	10	HS356B7	AL031714 Human DNA
42	230	15.8	202004	11	H0AF001549	AF001549 Human DNA
43	229.8	15.8	143747	9	HS370M22	Z82206 Human DNA
44	229.8	15.8	177236	34	AC005143	AC005143 Homo sapi
45	229.6	15.7	163871	9	HS1158E12	AL031584 Human DNA

## ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	NID	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES
1	AF135023	4740 bp	DNA	PRI	20-APR-1999		human.	Homo sapiens	Yousef,G.M., Luo,L.Y. and Diamandis,E.P.	1 (bases 1 to 4740)	Identification of novel human kallikrein-1-like genes on chromosome 19q13.3-q13.4	Biochem. Biophys. Res. Commun. (1999) In press	Yousef,G.M., Luo,L.Y. and Diamandis,E.P.	Direct Submission	Submitted (13-MAR-1999) Pathology and Laboratory Medicine, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada	Location/Qualifiers	

```

source
1. .4740
/organism="Homo sapiens"
/db.xref="taxon:9606"
/chromosome="19"
/map="19q13.3-q13.4"
join<2263. .2425,2847. .3097,3181. .3317,4588. .>4740)
/product="kallikrein-like protein 1 KLK-L1"
join<2263. .2425,2847. .3097,3181. .3317,4588. .4740)
/codon.start=3
/product="kallikrein-like protein 1 KLK-L1"
/protein.id="A02642.1"
/db.xref="PIR:G4589273"
/db.xref="GI:4589273"
/translation="SLVSGSCSOIINGEDCSPHSQPMOALVMEHELFCSGVLPVH
VLSAHCDFNSYITIGLHLEADQEPGSMVLAISVHPPEYNRPLANDMLIKLD
ESVSDIRISITISOCPTAGNSCLVSGMGLARMPYVQCINVSIVSEVCSKL
YDPLVPSMFCAGGGGDDKDCNSGSGGLICNGYLGGLVSGKAPCGOVPGVYTN
LCKTEWLEKTVQAS"
BASE COUNT 1106 a 1275 c 1308 g 1051 t
ORIGIN

Query Match 96.7% Score 1411.4; DB 42; Length 4740;
Best Local Similarity 99.6% Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

33 CAGAGCTCTACACATCGGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGG 92
2842 CACAGCTCTACACATCGGGGCTGGGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGG 2901
93 AGCCAGATGGTGAGGCGCAGCTCTCCGTAAGGCGACCCAGAGTAACAAGACCTTGCTC 152
2902 AGCCAGATGGTGAGGCGCAGCTCTCCGTAAGGCGACCCAGAGTAACAAGACCTTGCTC 2961
153 GCCTAAGACTCTAGTCTCATCAAGTTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGG 212
2962 GCCTAAGACTCTAGTCTCATCAAGTTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGG 3021
213 AGCATCAGATTGCTTCGAGTGCCTTACCGGGGGGAACTTGCTCGTTCTGGCTGG 272
3022 AGCATCAGATTGCTTCGAGTGCCTTACCGGGGGGAACTTGCTCGTTCTGGCTGG 3081
273 GGTCTGCTGGCGAAGCGTGAAGTCTCAGGGGTGTGTCTGCTCCCTCTTCAAGAGAGTCTCT 332
3082 GGTCTGCTGGCGAAGCGTGAAGTCTCAGGGGTGTGTCTGCTCCCTCTTCAAGAGAGTCTCT 3141
333 GCCCACTCGCGGGGGCTGAACCCAGAGCTTGCCTCCAGGCGAGATGCTTACCCTGCTGC 392
3142 GCCCACTCGCGGGGGCTGAACCCAGAGCTTGCCTCCAGGCGAGATGCTTACCCTGCTGC 3201
393 AGTGGTGAAGTGTGGGTGTGTGANGAGTCTGANTAGCTATGACCCGCTGT 452
3202 AGTGGTGAAGTGTGGGTGTGTGANGAGTCTGANTAGCTATGACCCGCTGT 3261
453 ACCACCCCANCATGTTCTCGCGCGCGGCGAGGCGCAAGACAGAGAGTCTCTGCAACGTGA 512
3262 ACCACCCCANCATGTTCTCGCGCGCGGCGAGGCGCAAGACAGAGAGTCTCTGCAACGTGA 3321
513 GAGAGGGGAAAAGGGGAGGCGAGCTCAAGGGAAGGTTGGAGAGGGGAGACAGAGAC 572
3322 GAGAGGGGAAAAGGGGAGGCGAGCTCAAGGGAAGGTTGGAGAGGGGAGACAGAGAC 3381
573 ACACAGGGGCGCATGGCGATGACAGATGAGAGACACACAGGGAGACAGAGCAACT 632
3382 ACACAGGGGCGCATGGCGATGACAGATGAGAGACACACAGGGAGACAGAGCAACT 3441
633 AGAGAGAGAACTGAGAGAAACAGAGAAATTAACACAGAGAAATTAAGAGAGCAAGAGAG 692
3442 AGAGAGAGAACTGAGAGAAACAGAGAAATTAACACAGAGAAATTAAGAGAGCAAGAGAG 3501
693 AGAGAGAGAAACAGACATGGGAGGCGAGAAACACACACACATAGAAATGCAGTTGACC 752
3502 AGAGAGAGAAACAGACATGGGAGGCGAGAAACACACACACATAGAAATGCAGTTGACC 3561

```

```

753 TTCACAGCATGGGGGCTGAGGGCGGTGACCTCCACCAATAGAAAATCCTTATAC 812
3562 TTCACAGCATGGGGGCTGAGGGCGGTGACCTCCACCAATAGAAAATCCTTATAC 3621
813 TTTTACTCCCAAAAACCTGACTGAAATAGCTTACTGTGAGGGAGCCTTACCAAT 872
3622 TTTTACTCCCAAAAACCTGACTGAAATAGCTTACTGTGAGGGAGCCTTACCAAT 3681
873 AACATAAATAGTCGATTTATGATACGTTTATGATTCATGATATACCTTTGTCGAT 932
3682 AACATAAATAGTCGATTTATGATACGTTTATGATTCATGATATACCTTTGTCGAT 3741
933 TTTTGAATTTCTAAGCTACACAGTTGCTGTGAAATTTTAAATTTGGAACCT 992
3742 TTTTGAATTTCTAAGCTACACAGTTGCTGTGAAATTTTAAATTTGGAACCT 3801
993 CCTAAATTTTTCGATGTTTATGAAAAATCAAGTAAATGAGTGTGCA-TT 1051
3802 CCTAAATTTTTCGATGTTTATGAAAAATCAAGTAAATGAGTGTGCA-TT 3861
1053 CAACACAGGTTGTTCAAGGTTCAACTGTACCCAGGGAACAGTACACAGATTC 1111
3862 CAACACAGGTTGTTCAAGGTTCAACTGTACCCAGGGAACAGTACACAGATTC 3921
1112 TAGAGTGAACACAGAGAGAAACAGAAAAATCAAGCTGTACAAAGAGCTGGCAGG 1171
3922 TAGAGTGAACACAGAGAGAAACAGAAAAATCAAGCTGTACAAAGAGCTGGCAGG 3981
1172 GTGGCTATGCTGTAAATCCAGACCTTTGGAGGCGAGGCGAGATCACTTGAGGTA 1231
3982 GTGGCTATGCTGTAAATCCAGACCTTTGGAGGCGAGGCGAGGAGATCACTTGAGGTA 4041
1232 AGGATTTCAAGACAGCCTTGCCCAAAATGCTGAATCTGTCTACTTAAATACAAA 1291
4042 AGGATTTCAAGACAGCCTTGCCCAAAATGCTGAATCTGTCTACTTAAATACAAA 4101
1292 GTTACTGATATGTTGGGCGGCGCTGTAAATCCAGCTACTTGGAGGCTGGAGCGA 1351
4102 GTTACTGATATGTTGGGCGGCGCTGTAAATCCAGCTACTTGGAGGCTGGAGCGA 4161
1352 GAATTCCTGAATATGAGGCGAGAGGTTGAAGTGAATGATCACACCACTATCTCC 1411
4162 GAATTCCTGAATATGAGGCGAGAGGTTGAAGTGAATGATCACACCACTATCTCC 4221
1412 AGCTGGGCAACAGTAAGCTGTCTCAAAAAAATTTTTTTTTT 1459
4222 AGCTGGGCAACAGTAAGCTGTCTCAAAAAAATTTTTTTTTT 4269

RESULT 2
AF148532 4385 bp DNA 09-JUN-1999
LOCUS AF148532 Homo sapiens kallikrein 4 (KLK4) gene, complete cds.
DEFINITION AF148532
ACCESSION AF148532
NID 95020095
VERSION AF148532.1 GI:5020095
KEYWORDS
SOURCE
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4385)
AUTHORS Stephenson,S.A., Verity,K., Ashworth,L. and Clements,J.A.
TITLE Localization of a new prostate specific antigen-related serine
proteinase gene, KLK4, is evidence for an expanded human kallikrein
(KLK) gene family cluster on chromosome 19q13.3-13.4
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4385)
AUTHORS Stephenson,S.A., Ashworth,L. and Clements,J.A.
TITLE Direct Submission
JOURNAL Submitted (03-MAY-1999) School of Life Science, Queensland
University of Technology, George Street, Brisbane, Queensland 4001,

```



```

FEATURES             Australia
Source                Location/Qualifiers
                     1..4385
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /map="19q13.3-q13.4"
                     /join("<551..611,1875..2037,2459..2709,2793..2929,
mRNA                  4200..>4352)
                     /gene="KLK4"
                     /product="Kallikrein 4"
                     <551..>4352
gene                  /gene="KLK4"
                     /join("551..611,1875..2037,2459..2709,2793..2929,4200..4352)
CDS                   /gene="KLK4"
                     /function="serine protease"
                     /codon_start=1
                     /product="Kallikrein 4"
                     /protein_id="A038019.1"
                     /db_xref="pid:95020096"
                     /db_xref="gi:5020096"
                     /translation="MATAPNPMGFLGLGLVAGSLVSSCSQIINGEDCSPHSOPM
OALVMENELFCGVLVHPQWVLSAHCFO NSTITGLHSLDEDPGSMVEASLS
VRHEYNRLANDLMLIKLDESVEEDTIRISLISIGCTAGNSCLVSGPMLANGR
MPYVLOCVNVSVSEVCSKLYDPLVHPMFCAAGGGDDKDCNDBSGGFLICNGLYLQ
GLVSEGAFCGQVGPVNTNLCKTFEMIKTYQAS"
BASE COUNT           1013 a 1184 c 1192 g 996 t
ORIGIN
Query Match          96.7% Score 1411.4; DB 42; Length 4385;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1423; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 33 CAGAGCTCTACACCATCGGGCTGGCCTGCACAGTCTTGAGGCCAACAAGCCAGGG 92
    |||
    |||
    |||
Db 2454 CACAGCTCTACACCATCGGGCTGGCCTGCACAGTCTTGAGGCCAACAAGCCAGGG 2513
QY 93 AGCCAGATGGGAGGAGGAGCCCTCGCTGAGGACCCAGAGTCAACAAGACCTTGCTC 152
    |||
    |||
    |||
Db 2514 AGCCAGATGGGAGGAGGAGCCCTCGCTGAGGACCCAGAGTCAACAAGACCTTGCTC 2573
QY 153 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGG 212
    |||
    |||
    |||
Db 2574 GCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGTCTGACACCATCCGG 2633
QY 213 AGCATCAGCATGCTTGGCACTGCCCTACCGGGGGAATCTTGCCTGTTTGGCTGG 272
    |||
    |||
    |||
Db 2634 AGCATCAGCATGCTTGGCACTGCCCTACCGGGGGAATCTTGCCTGTTTGGCTGG 2693
QY 273 GGTCGTGGGGAAGGAGCTCAGGGGTGTGTCTGCGCCCTTTCAAGAGAGTCCCTCT 332
    |||
    |||
    |||
Db 2694 GGTCGTGGGGAAGGAGCTCAGGGGTGTGTCTGCGCCCTTTCAAGAGAGTCCCTCT 2753
QY 333 GCCCAGTCCGGGGGCTGACCCAGAGCTGTGCTCCAGGAGAGATGCCCTACCTGTGC 392
    |||
    |||
    |||
Db 2754 GCCCAGTCCGGGGGCTGACCCAGAGCTGTGCTCCAGGAGAGATGCCCTACCTGTGC 2813
QY 393 AGTCGCTGAACGTGTGGTGTGTCTGANGAGTCTGCANTTAAGCTATGACCCGCTGT 452
    |||
    |||
    |||
Db 2814 AGTCGCTGAACGTGTGGTGTGTCTGANGAGTCTGCANTTAAGCTATGACCCGCTGT 2873
QY 453 ACCACCCCAACATGTTGTCGGCGCGGAGGAGGCAAGCCAAAGAGATCCCTGCAACGTGA 512
    |||
    |||
    |||
Db 2874 ACCACCCCAACATGTTGTCGGCGCGGAGGAGGCAAGCCAAAGAGATCCCTGCAACGTGA 2933
QY 513 GAGAGGGGAAAGGGGAGGAGGAGGAGCTCAGGGAAGGGTGGAGAAGGGGAGAGAGAGAC 572
    |||
    |||
    |||
Db 2934 GAGAGGGGAAAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2993
QY 573 ACACAGGGGCGCATGGCAGATGCAAGAGATGAGAGACACACAGGAGAGACAGTGAACAAT 632
    |||
    |||
    |||
Db 2994 ACACAGGGGCGCATGGCAGATGCAAGAGATGAGAGACACACAGGAGAGACAGTGAACAAT 3053

```

```

QY 633 AGAGAGAAAGTGAAGAAACAGAGAAATTAACACAGAAATTAAGAGAAAGAAAG 692
    |||
    |||
    |||
Db 3054 AGAGAGAAAGTGAAGAAACAGAGAAATTAACACAGAAATTAAGAGAAAGAAAG 3113
QY 693 AGAGAAACAGAAACAGACATGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 752
    |||
    |||
    |||
Db 3114 AGAGAAACAGAAACAGACATGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3173
QY 753 TTCCACAGCATGGGGGCTGAGGGCGGTGACCTCCACCCCAATAGAAATCCTTTATAC 812
    |||
    |||
    |||
Db 3174 TTCCACAGCATGGGGGCTGAGGGCGGTGACCTCCACCCCAATAGAAATCCTTTATAC 3233
QY 813 TTTTGACTCCCAAAACCTGACTAGAAATAGCTTACTGTGAGGGGAGCCTTACCAAT 872
    |||
    |||
    |||
Db 3234 TTTTGACTCCCAAAACCTGACTAGAAATAGCTTACTGTGAGGGGAGCCTTACCAAT 3293
QY 873 AACATAATAGTGAATTTATGATACGTTTATGATCATGATATACCTTTGTTGGAAT 932
    |||
    |||
    |||
Db 3294 AACATAATAGTGAATTTATGATACGTTTATGATCATGATATACCTTTGTTGGAAT 3353
QY 933 TTTTGTATATTTCTAAGCTACACAGTTCGTCTGTAATTTTTTAATTGTTCAACTCT 992
    |||
    |||
    |||
Db 3354 TTTTGTATATTTCTAAGCTACACAGTTCGTCTGTAATTTTTTAATTGTTCAACTCT 3413
QY 993 CCTAAATTTTCTGATGTGTTTATTAAGAAATCCAAATATAGTGGACTGTGCA -TT 1051
    |||
    |||
    |||
Db 3414 CCTAAATTTTCTGATGTGTTTATTAAGAAATCCAAATATAGTGGACTGTGCA -TT 3473
QY 1052 CAACACAGGTTGTTCAAGGTCACACTGTGTACCCAGAGGAGAAACAGTGACAGATTCA 1111
    |||
    |||
    |||
Db 3474 CAACACAGGTTGTTCAAGGTCACACTGTGTGTACCCAGAGGAGAAACAGTGACAGATTCA 3533
QY 1112 TAGAGTGAAACACAGAAAGAGAAACAGAAATCAAGACTTCAACAAAGAGCTGGCAGG 1171
    |||
    |||
    |||
Db 3534 TAGAGTGAAACACAGAAAGAGAAACAGAAATCAAGACTTCAACAAAGAGCTGGCAGG 3593
QY 1172 GTGGCTCATGCTGTATATCCAGACTTTGGAGGCGAGGACGAGCATCTTGAGGTA 1231
    |||
    |||
    |||
Db 3594 GTGGCTCATGCTGTATATCCAGACTTTGGAGGCGAGGACGAGCATCTTGAGGTA 3653
QY 1232 AGGAGTTCAGAGACAGCCTGGCCAAATGTTGAATCTCTGTACTAAATAATACAAA 1291
    |||
    |||
    |||
Db 3654 AGGAGTTCAGAGACAGCCTGGCCAAATGTTGAATCTCTGTACTAAATAATACAAA 3713
QY 1292 GTTAGCTGATATGTTGGAGGCGCCTGTATATCCAGACTTGGAGGCTGAGGAGGA 1351
    |||
    |||
    |||
Db 3714 GTTAGCTGATATGTTGGAGGCGCCTGTATATCCAGACTTGGAGGCTGAGGAGGA 3773
QY 1352 GAATTCGTTAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1411
    |||
    |||
    |||
Db 3774 GAATTCGTTAATATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3833
QY 1412 AGCTGGGGAACAGATGAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAAA 1459
    |||
    |||
    |||
Db 3834 AGCTGGGGAACAGATGAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAAA 3881

```

```

RESULT 3
AF113141
LOCUS      AF113141      5900 bp      DNA
DEFINITION Homo sapiens serine protease gene, complete cds.
ACCESSION AF113141
VERSION    94512031
KEYWORDS   AF113141.1 GI:4512031
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 5900)
            Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and
            Wang,X.

```

TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression  
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)  
JOURNAL 99179024  
MEDLINE 2 (bases 1 to 5900)  
REFERENCE Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.  
AUTHORS  
TITLE Direct Submission  
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE, Bothell, WA 98021, USA  
FEATURES  
source  
Location/Qualifiers  
1. 5900  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13"  
join(<1411, .1471, 2735, .2897, 3319, .3569, 3653, .3789, 5062, .5817)  
/product="serine protease prostate"  
1411, 1471  
/number=1  
join(1411, .1471, 2735, .2897, 3319, .3569, 3653, .3789, 5062, .5214)  
/note="androgen regulated"  
/codon\_start=1  
/product="serine protease prostate"  
/protein\_id="AND21581.1"  
/db\_xref="PDB:4512032"  
/db\_xref="GI:4512032"  
/translation="MATAGNPWGFWFLYLILGAVSLVSGSCQIINGEDCSPHSQPM QALVMENELFCSGVLVHPQVLSAHCPCQNSYTIIGLHSLAEQDEGSDQWVASLS VHPPEYRRLPLNDMLIKLDESVESEDTRISITIASOCPNAGNSCLVSGWELLANGR MPTVLCQVNVSVSSEVSKLYDPLVHPMPEFAGGSHDKDSCNDSGPLICNRYLQ GLVSGKAPCGGVGVGVYTNLCFTEWLEKTVQNS"  
2735, .2897  
/number=2  
3319, .3569  
/number=3  
3653, .3789  
/number=4  
5062, .5817  
/number=5  
BASE COUNT 1417 a 1643 c 1569 g 1271 t  
ORIGIN  
Query Match 96.3%; Score 1405; DB 42; Length 5900;  
Best Local Similarity 99.4%; Pred. No. 0;  
Matches 1419; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

DB 3614 GCCCAGTCGGGGGGCTGACCAGAGCTCTGCTCCAGCGAATGCTACCGTCTGC 3673  
QY 393 AGTGGGTGAACCTGTGCTGTCTGANGAGTCTTGCAANTAGCTATACCCCGCTG 452  
DB 3674 AGTGGGTGAACCTGTGCTGTCTGANGAGTCTTGCAANTAGCTATACCCCGCTG 3733  
QY 453 ACCACCCCMCATGTTCTGCGCGCGGAGGCAAGACCAAGAGACTCTGCAACGTGA 512  
DB 3734 ACCACCCCMCATGTTCTGCGCGCGGAGGCAAGACCAAGAGACTCTGCAACGTGA 3793  
QY 513 GAGAGGGGAAAGGGAGGCGACGCTCAGGGAAGGTTGAGAAAGGGGAGACAGAGAC 572  
DB 3794 GAGAGGGGAAAGGGAGGCGACGCTCAGGGAAGGTTGAGAAAGGGGAGACAGAGAC 3853  
QY 573 AACAGGGCGCGATGGCCAGATGCAAGATGAGAGACACAGGAGACAGTGAACACT 632  
DB 3854 AACAGGGCGCGATGGCCAGATGCAAGATGAGAGACACAGGAGACAGTGAACACT 3913  
QY 633 AGAGAGAGAACTGAGAGAAACAGAGAAATAACACAGGAATTAAGAGACAAAGAG 692  
DB 3914 AGAGAGAGAACTGAGAGAAACAGAGAAATAACACAGGAATTAAGAGACAAAGAG 3973  
QY 693 AGAGAGAGAAACAGAGATGGGAGGAGCAAAACACACACATAGAAATGACGTTGACC 752  
DB 3974 AGAGAGAGAAACAGAGATGGGAGGAGCAAAACACACACATAGAAATGACGTTGACC 4033  
QY 753 TTCCAACAGCAGAGGGGCTGAGGGGGGTGACCTCCACCAATAGAAATCCTCTTATAC 812  
DB 4034 TTCCAACAGCAGAGGGGCTGAGGGGGGTGACCTCCACCAACAGAAATCCTCTTATAC 4093  
QY 813 TTTTACTCTCCCAAAACCTGACTAGAAATAGCTACTGTTGAGGGAGGCTTACCAAT 872  
DB 4094 TTTTACTCTCCCAAAACCTGACTAGAAATAGCTACTGTTGAGGGAGGCTTACCAAT 4153  
QY 873 AACATTAATAGTCGATTTATGCAATGTTTATGCAATTCATGATATACCTTTGTTGAT 932  
DB 4154 AACATTAATAGTCGATTTATGCAATGTTTATGCAATTCATGATATACCTTTGTTGAT 4213  
QY 933 TTTTGTATATTTCTTAAGCTACACATGCTGTGTAATTTTTTAAATGTTGCAACT 992  
DB 4214 TTTTGTATATTTCTTAAGCTACACATGCTGTGTAATTTTTTAAATGTTGCAACT 4273  
QY 993 CCTAAATTTTCTATATGTTTATGAAAAATCAAGTAAATGAGTGTGCA-TT 1051  
DB 4274 CCTAAATTTTCTATATGTTTATGAAAAATCAAGTAAATGAGTGTGCAACTT 4333  
QY 1052 CAACACAGGTTGTTCAAGGTTCAACTGTGTACCCAGAGGAAACAGTGACACAGATTCA 1111  
DB 4334 CAACACAGGTTGTTCAAGGTTCAACTGTGTACCCAGAGGAAACAGTGACACAGATTCA 4393  
QY 1112 TAGAGGTGAACACAGAAAGAAACAGAAATCAAGCTCTACAAAGAGCTGGCAGG 1171  
DB 4394 TAGAGGTGAACACAGAAAGAAACAGAAATCAAGCTCTACAAAGAGCTGGCAGG 4453  
QY 1172 GTGGCTCATGCTGTAAATCCAGCACTTTGGGAGCGAGGACAGATCACTTGAGGTA 1231  
DB 4454 GTGGCTCATGCTGTAAATCCAGCACTTTGGGAGCGAGGACAGATCACTTGAGGTA 4513  
QY 1232 AGGAGTTCAAGCACAGCTGGCCAAATGTGAATCTGTCTGTACTTAAATCAAAA 1291  
DB 4514 AGGAGTTCAAGCACAGCTGGCCAAATGTGAATCTGTCTGTACTTAAATCAAAA 4573  
QY 1292 GTTACTGGATATGTTGGCAGGCGCTTAATCCAGTACTTGGGAGCGGAGGACAGA 1351  
DB 4574 GTTACTGGATATGTTGGCAGGCGCTTAATCCAGTACTTGGGAGCGGAGGACAGA 1411  
QY 1352 GAATGCTTGAATATGAGAGGAGGAGTGAAGTGAAGTGAATCAACCACTATATCC 1411  
DB 4634 GAATGCTTGAATATGAGAGGAGGAGTGAAGTGAAGTGAATCAACCACTATATCC 4693  
QY 1412 AGCTGGGCAACAGAGTAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459  
DB 4694 AGCTGGGCAACAGAGTAAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 4741

```

RESULT 4
AF113140 1347 bp mRNA PRI 25-MAR-1999
LOCUS Homo sapiens serine protease mRNA, complete cds.
DEFINITION AF113140
ACCESSION 94512029
VERSION AF113140.1 GI:4512029
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1347)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and
Wang,K.
TITLE Molecular cloning and characterization of prostate, an
androgen-regulated serine protease with prostate-restricted
expression
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)
MEDLINE 99179024
REFERENCE 2 (bases 1 to 1347)
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and
Wang,K.
TITLE Direct Submision
JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,
Bothell, WA 98021, USA
FEATURES
source
1. .1347
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13"
1. .765
/note="androgen regulated"
/codon_start=1
/product="serine protease"
/protein_id="A021580.1"
/db_xref="PIR:G4512030"
/db_xref="GI:4512030"
/translacion="MATAGNPMWFLGILVAGSLVSGSQIINGEDCPHSQPM
QALVLENELFCGVLVHPQWVLAHCFQNSYIGLHSLHSDADPGQOMVEASLS
VRRPEYRNPLANDMLIKDESSESDDTSISIASOCTAGNSGLVSGWGLANGR
MPTVLOCVNVSVSEVSCSKLYDPLVHPSMFCAGSGHDDQDSCNDSGGLINGTILQ
GLVSRAPCGGVGVPGVYTNLCKFTWELKTVQAS"
polya_site 269 a 489 c 334 g 255 t
BASE COUNT 269 a 489 c 334 g 255 t
ORIGIN
Query Match 20.3%; Score 296.4; DB 42; Length 1347;
Best Local Similarity 81.0%; Pred. No. 2e-59;
Matches 393; Conservative 0; Mismatches 9; Indels 83; Gaps 1;
25 AGTGAGTGCAGAGTCTCTACACCATCGGGTGGGCGCTGCACAGCTTGAGGCCACCAAG 84
|||||
212 ACTGTTCCAGAACTCTCTACCATCGGGTGGGCGCTGCACAGCTTGAGGCCACCAAG 271
85 AGCCAGGAGCAGATGATGAGAGCCAGCTCTCCGTACGACCCAGAGTACAAGAC 144
|||||
272 AGCCAGGAGCAGATGATGAGAGCCAGCTCTCCGTACGACCCAGAGTACAAGAC 331
145 CTTTGCTGCTACGACCTCATGCTCAATGGAAGAAATCGGTCCGAGTGTGACA 204
|||||
332 CTTTGCTGCTACGACCTCATGCTCAATGGAAGAAATCGGTCCGAGTGTGACA 391
205 CCATCGGAGAGATGATGAGAGCCAGCTCTCCGTACGACCCAGAGTACAAGAC 264
|||||
392 CCATCGGAGAGATGATGAGAGCCAGCTCTCCGTACGACCCAGAGTACAAGAC 451
265 CTGGCTGGGCTGCTGGCGAAGCGTACAGGGTGTGTCTGCTCCCTCTCAAGA 324
|||||

```

```

Db 452 CTGGCTGGGCTGCTGGCGAAG----- 475
QY 325 GGTCTCTGCCAGTGGCGGGGCTGACCCAGAGCTCTGCTCCAGCGAATGCTTAC 384
Db 475 -----GGCGAATGCTTAC 488
QY 385 CGTGCTCAGTGGTGAACGTGCGGTGTGTGANGAGTGTGCAATPAAGCTTATGA 444
|||||
Db 489 CGTGCTCAGTGGTGAACGTGCGGTGTGTGANGAGTGTGCAATPAAGCTTATGA 548
445 CCCGCTGTACCCACCCANCATGTTTGGCGGCGGAGGAGGACAGAGACTTCTG 504
|||||
Db 549 CCGCTGTACACCCAGCATGTTTGGCGGCGGAGGAGGACAGAGACTTCTG 608
QY 505 CAACG 509
|||||
Db 609 CAACG 613

RESULT 5
HS310013 166723 bp DNA HTG 10-JUN-1999
LOCUS Homo sapiens chromosome 20 clone 310013, WORKING DRAFT SEQUENCE, In
DEFINITION unordered pieces.
ACCESSION AL031658
NID 94902630
VERSION AL031658.10 GI:4902630
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166723)
REFERENCE 1
AUTHORS Kay,M.
TITLE Direct Submision
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humber@esanger.ac.uk Clone requests: clonerequests@esanger.ac.uk
On May 27, 1999 this sequence version replaced gi:4741461.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Unfinished: dj310013 Contig_ID: 00252 acc-AL031658
length: 1113 bp Unfinished: dj310013 Contig_ID: 00912
acc-AL031658 length: 162976 bp unfinished: dj310013 Contig_ID:
02100 acc-AL031658 length: 1034 bp
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
source
1. 166723
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="20"
/clone="310013"
BASE COUNT 42104 a 40443 c 39486 g 43089 t 1601 others
ORIGIN
Query Match 17.0%; Score 248.2; DB 34; Length 166723;
Best Local Similarity 84.3%; Pred. No. 4.8e-48;
Matches 291; Conservative 0; Mismatches 53; Indels 1; Gaps 1;
1113 AGAGGTAAACAGAGAAAGAAATCAAGACTTACAAAGAGCTGCGCAGGG 1172
|||||
Db 93064 AATTTTCAACATCTCTTAAATAAGACTTCAAAACATCGTAGTGGCATGG 93123
1173 TGCTCATGCTGTAAATCCAGCACTTGGGAGG-CGAGGACAGCAATCACTGAGGTA 1231
|||||

```

Db 93124 TGGCTCATGCTGTAAATCCAGCATTGTGGAGCCGAGGTGGGAAGATCATTGAGGTC 93183  
 Oy 1232 AGAGTTCAAGACAGCTGGCCAAATGGTAAATCTGTCTGTACTAAATAACAAAA 1291  
 Db 93184 AGGAGTTCAAGACAGCTGGCCAAATGGTAAATCTGTCTGTACTAAATAACAAAA 93243  
 Oy 1292 GTTACGTGATATGTGGCAGGCGCCTGTATCCAGTACTTGGGAGGCTGAGCAGGA 1351  
 Db 93244 ATTAGCTGGGCATGTGGCAGCAGCTGTAGTCCAGCTACTCAGAGGCTGAGCAGGA 93303  
 Oy 1352 GAATTCCTTGAATATGGGAGCAGAGTGAATGAGATCAGCAGCATATCTCC 1411  
 Db 93304 GAATTCCTTGAATATGGGAGCAGAGTGAATGAGATCAGCAGCATATCTCC 93363  
 Oy 1412 AGCTGGGCACAGTAGTACTGTCTCAAAAAA 1456  
 Db 93364 AGCTGGGCACAGTAGTACTGTCTCAAAAAA 93408

RESULT 6  
 LOCUS HSAC002066 151770 bp DNA PRI 09-MAY-1997  
 DEFINITION Human BAC clone RG030H15 from 7q31, complete sequence.  
 ACCESSION AC002066  
 NID 92076721  
 VERSION AC002066.1 GI:2076721  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homidae; Homo.  
 REFERENCE 1 (bases 1 to 151770)  
 AUTHORS Dante,M, Mamsley,P and Keppler,D.  
 JOURNAL The sequence of H. sapiens BAC clone RG030H15  
 REFERENCE 2 (bases 1 to 151770)  
 AUTHORS Waterston,R.  
 JOURNAL Direct Submission  
 COMMENT Submitted (09-MAY-1997)  
 Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63108, USA  
 http://genome.wustl.edu/gsc  
 e-mail: saplens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7> or send an E-mail to [egreen@nhgri.nih.gov](mailto:egreen@nhgri.nih.gov)

SOURCE INFORMATION:  
 This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
 VECTOR: pBEO  
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
 Actual start of this clone is at base position 1 of H\_RG030H15.  
 Actual end is at 151770 of H\_RG030H15. The orientation of this clone is unknown.

This clone contains STS SWSS965.

FEATURES	Location/Qualifiers
source	1..151770
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/clone="RG030H15"
	/clone_1lb="CTB-978SK-B"
	/map="7q31"
repeat_region	complement(671..690)
	/rpt_family="L1"
repeat_region	2544..2593
	/rpt_family="L1"
repeat_region	complement(2759..2781)
	/rpt_family="L1"
repeat_region	complement(4347..4468)
	/rpt_family="L1"
repeat_region	complement(5317..5615)
	/rpt_family="ALU"
repeat_region	complement(6063..6314)
	/rpt_family="ALU"
repeat_region	complement(7577..8152)
	/rpt_family="L1"
repeat_region	complement(8176..8233)
	/rpt_family="L1"
repeat_region	complement(8706..9002)
	/rpt_family="ALU"
repeat_region	complement(9318..9390)
	/rpt_family="ALU"
repeat_region	complement(9403..9434)
	/rpt_family="L1"
repeat_region	complement(9481..9510)
	/rpt_family="L1"
repeat_region	9513..9789
	/rpt_family="ALU"
repeat_region	10325..10770
	/rpt_family="L1"
misc_feature	12726..13009
	/note="similar to human EST D45516 (NID:9662470)"
repeat_region	14161..15579
	/rpt_family="L1"
repeat_region	15804..15902
	/rpt_family="L1"
repeat_region	17101..17218
	/rpt_family="ALU"
repeat_region	complement(117462..17485)
	/rpt_family="L1"
repeat_region	complement(17519..17675)
	/rpt_family="ALU"
repeat_region	complement(18866..19055)
	/rpt_family="L1"
repeat_region	complement(19068..20712)
	/rpt_family="L1"
repeat_region	complement(20933..21001)
	/rpt_family="L1"
repeat_region	complement(21021..21785)
	/rpt_family="L1"
repeat_region	complement(21919..21943)
	/rpt_family="L1"
repeat_region	23706..24283
	/rpt_family="L1"
repeat_region	25651..25842
	/rpt_family="ALU"
repeat_region	27166..27287
	/rpt_family="ALU"
repeat_region	27292..27794
	/rpt_family="L1"
repeat_region	complement(27974..28348)



AC007728/C	LOCUS	AC007728	166663 bp	DNA	HTG	05-JUN-1999
DEFINITION		Homn sapiens chromosome 16 clone 327F22, WORKING DRAFT SEQUENCE, 5 ordered pieces.				
ACCESSION		AC007728				
NID		95001450				
VERSION		AC007728.1	GI:5001450			
KEYWORDS		HTG; HTGS_PHASE2.				
SOURCE		human.				
ORGANISM		Homn sapiens				
REFERENCE		Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS		1 (bases 1 to 166663)				
		Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,D., White,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R., Han,C. and Deaven,L.				
TITLE		Sequencing of Human Chromosome 16q12				
JOURNAL		Unpublished				
REFERENCE		2 (bases 1 to 166663)				
AUTHORS		Rice,D.O.				
TITLE		Large Scale Sequence Analysis and Annotation with the Sequence Comparison Analysis (SCAN) System				
JOURNAL		Unpublished				
AUTHORS		3 (bases 1 to 166663)				
		Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesmer,J., Meincke,L., Longmire,D., White,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R., McMurry,K., Han,C. and Deaven,L.				
TITLE		Direct Submission				
JOURNAL		Submitted (05-JUN-1999) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA				
COMMENT		* NOTE: This is a 'working draft' sequence. It currently				
		* consists of 5 contigs. Gaps between the contigs				
		* are represented as runs of N. The order of the pieces				
		* is believed to be correct as given, however the sizes				
		* of the gaps between them are based on estimates that have				
		* provided by the submittor.				
		* This sequence will be replaced				
		* by the finished sequence as soon as it is available and				
		* the accession number will be preserved				
		1 22988: contig of 22988 bp in length				
		* gap of unknown length				
		* 22989 109379: contig of 86391 bp in length				
		* gap of unknown length				
		* 109380 154579: contig of 45300 bp in length				
		* gap of unknown length				
		* 154580 156226: contig of 1647 bp in length				
		* gap of unknown length				
		* 156227 166663: contig of 10437 bp in length.				
FEATURES		Location/Qualifiers				
source		1. 166663				
		/organism="Homn sapiens"				
		/db_xref="taxon:9606"				
		/clone="327F22"				
		/chromosome="16"				
BASE COUNT		5110 a 37036 c 34518 g 43999 t				
ORIGIN						
Query Match		16.4%: Score 240; DB 35; Length 166663;				
Best Local Similarity		83.5%: Pred. No. 3,9e-46;				
Matches 284; Conservative 0; Mismatches 55; Indels 1; Gaps 1						
QY 1121 AACACGACGAGAAACAGGAAAAATCAAGACTCTACAAAGAGCGTGGGACGAGGTGCTCAT 1180						
Db 111668 AACGACACCAAAACACCATCAACAGAAAGGACACAAAATATGCGCGGACACATTTGGCTCAT 111609						
QY 1181 GCCTGTAATCCACGACCTTTGGAGGC -GAGGACGACGACATCTACTTGGAGTAAGGATTC 1239						

D	111608	GCCCTATATCCAGACACTTTGGGAGGCGTGAAGCAGGCAAGCATCACTTGAGGTGCAGAGCTT	111549
OY	1240	AAGACCAGCCTGTCGCCAAATAATGCTGAATTCCTGTCGTACTTAATAATACAAAATTAGCTG	1299
D	111548	GAGCACACCTCTGCTCAAAATATGGTGAACCTCTCTCTCTAATAAATACAAAAATTAGCTG	111489
OY	1300	GATATGATGGGCGAGGCGCCTGTAAATCCAGACTACTTGGGAGGCTGAGCGCAGAGAATTGCT	1359
D	111488	GGTGTGTGTGGCAGGCACTGTATATCCAGCTACTTGGGAGGCTGAGCGCAGAGAAATTGCT	111429
OY	1360	TGAATATGGGAGGCGCAGAGCTGAAATGAGTTGAGATCACACCACTACTACTCAGCTGGGG	1419
D	111428	TGAACCTGGGAGGCGGAGGAGTTGCAGATGAGTGAGATCGCACATATGCACCTCAGGCTAGG	111369
OY	1420	CACACGAGTAGACTCTCTCTCAAAAAAAAAAAAAAAAAAAAAA	1459
D	111368	CACACAGCAAGACACTCTCTCAAAAACAACAAACAAACA	111329
RESULT	8		
LOCUS	AC002429		
DEFINITION	AC002429	Human BAC clone GS200K05 from 7q21-q22, complete sequence.	
ACCESSION	AC002429		
NID	92335067		
VERSION	AC002429.1	GI:2335067	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Cetarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 234053)		
JOURNAL	Graves,T, Duckels,G and Hawkins,M.		
REFERENCE	The sequence of H. sapiens BAC clone GS200K05		
AUTHORS	unpublished (1997)		
TITLE	2 (bases 1 to 234053)		
JOURNAL	Waterston,R.		
COMMENT	Direct Submission		
	Submitted (19-AUG-1997) Department of Genetics, Washington		
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA		
	SUBMITTED BY:		
	Genome Sequencing Center		
	Department of Genetics		
	Washington University		
	St. Louis MO 63108, USA		
	http://genome.wustl.edu/gsc		
	mailto:saplens@wustl.edu		
	NOTICE: This sequence may not represent the entire insert of this		
	clone. It may be shorter because we only sequence overlapping		
	clone sections once, or longer because we provide a small overlap		
	between neighboring data submissions.		
	This sequence was finished as follows unless otherwise noted:		
	all regions were double stranded or sequenced with an alternate		
	chemistry; an attempt was made to resolve all sequencing problems,		
	such as compressions and repeats; all regions were covered by		
	sequence from more than one subclone; and the assembly was		
	confirmed by restriction digest.		
	MAPPING INFORMATION:		
	The sequence of this clone was established as part of a mapping and		
	sequencing collaboration between the NHGRI Chromosome 7 Mapping		
	Project (Eric D. Green, Director), John D. McPherson in the		
	Department of Genetics (Washington University), and the Washington		
	University Genome Sequencing Center. For additional information		
	about the map position of this sequence, see		
	http://www.nhgri.nih.gov/DIR/GTB/CHR7 or send		
	mailto:egreen@nhgri.nih.gov		
	SOURCE INFORMATION:		
	This clone is from the first BAC library from Genome Systems, Inc.		
	(http://www.genomesystems.com).		

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

**MAPPING INFORMATION:**  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7> or send <mailto:egreen@nhgri.nih.gov>

**SOURCE INFORMATION:**  
This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).

Cell line: lymphoblastoid  
Haplotypes: two  
VECTOR: pBelOBAC  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
Actual start of this clone is at base position 1 of 1 GS200K05;  
actual end is at 234053 of GS200K05. The orientation of this clone  
is unknown.

This cosmid contains sts's SWS537 (NID:91222821) and SWS3746  
(NID:9151514).

## FEATURES

source	Location/Qualifiers
repeat_region	/rpt_family="L1"
repeat_region	/complement(21120..21414)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(21722..25632)
repeat_region	/rpt_family="L1"
repeat_region	/complement(23301..23274..23301)
repeat_region	/rpt_family="L1"
repeat_region	/complement(26636..26198..26636)
repeat_region	/rpt_family="L1"
repeat_region	/complement(27632..27653)
repeat_region	/rpt_family="L1"
repeat_region	/complement(27776..27803)
repeat_region	/rpt_family="L1"
repeat_region	/complement(28151..28127..28151)
repeat_region	/rpt_family="L1"
repeat_region	/complement(29826..29781..29826)
repeat_region	/rpt_family="L1"
repeat_region	/complement(30309..30326)
repeat_region	/rpt_family="L1"
repeat_region	/complement(30344..31472)
repeat_region	/rpt_family="L1"
repeat_region	/complement(31542..31829)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(32096..32133)
repeat_region	/rpt_family="L1"
repeat_region	/complement(32165..32729)
repeat_region	/rpt_family="L1"
repeat_region	/complement(32758..33562)
repeat_region	/rpt_family="L1"
repeat_region	/complement(34455..34486)
repeat_region	/rpt_family="L1"
repeat_region	/complement(34778..34492..34778)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(35297..35386)
repeat_region	/rpt_family="L1"
repeat_region	/complement(36329..35973..36329)
repeat_region	/rpt_family="L1"
repeat_region	/complement(36687..36981)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(37200..37162..37200)
repeat_region	/rpt_family="L1"
repeat_region	/complement(38651..38925)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(39644..39687)
repeat_region	/rpt_family="L1"
repeat_region	/complement(39688..39980)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(40197..40226)
repeat_region	/rpt_family="L1"
repeat_region	/complement(41184..41164..41184)
repeat_region	/rpt_family="L1"
repeat_region	/complement(41765..41716..41765)
repeat_region	/rpt_family="L1"
repeat_region	/complement(43753..43720..43753)
repeat_region	/rpt_family="L1"
repeat_region	/complement(45251..45223..45251)
repeat_region	/rpt_family="L1"
repeat_region	/complement(46066..46102)
repeat_region	/rpt_family="L1"
repeat_region	/complement(46105..46388)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(46748..47039)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(47958..47669..47958)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(49195..48675..49195)
repeat_region	/rpt_family="L1"
repeat_region	/complement(50408..49988..50277)
repeat_region	/rpt_family="ALU"
repeat_region	/complement(50408..50442)
repeat_region	/rpt_family="L1"
repeat_region	/complement(53772..53879)
repeat_region	/rpt_family="L1"

BASE COUNT	.	39502	a	33809	c	32673	g	36176	t	18	others
------------	---	-------	---	-------	---	-------	---	-------	---	----	--------

Double stranded (DS) coverage: 99.78  
DS or two chemistry coverage: 99.98



Single stranded regions: 1

# Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. Small fragments below a variable cutoff (approximately 400-600) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragment groups are separated by dashed lines.

BgIII		EcoRI		NsiI	
Map	Seq	Map	Seq	Map	Seq
1221.00	1225.00	777.00	753.00	2434.00	2433.00
9301.00	9214.00	2717.00	2724.00	2016.00	2016.00
2260.00	2247.00	6542.00	6545.00	958.00	925.00
542.00	540.00	2717.00	2671.00	15887.00	15655.00
2421.00	2383.00	3666.00	3619.00	1772.00	1786.00
4040.00	4054.00	2020.00	1958.00	2230.00	2233.00
1221.00	1208.00	3319.00	3354.00	5045.00	4995.00
1068.00	1047.00	7389.00	7370.00	527.00	518.00
1482.00	1487.00	847.00	842.00	1451.00	1421.00
9301.00	9202.00	3319.00	3252.00	958.00	945.00
5993.00	5469.00	10078.00	9931.00	5371.00	5359.00
4200.00	4199.00	5656.00	5592.00	1002.00	994.00
2816.00	2790.00	1896.00	1892.00	5045.00	4907.00
3863.00	3891.00	10078.00	10220.00	1072.00	1089.00
2616.00	2589.00	777.00	799.00	1002.00	999.00
7106.00	7198.00	6772.00	6793.00	3998.00	3967.00
8828.00	8842.00	2020.00	2027.00	1532.00	1507.00
5679.00	5807.00	2311.00	2325.00	2434.00	2427.00
1389.00	1385.00	11025.00	11100.00	1072.00	1080.00
5679.00	5685.00	8690.00	8656.00	2016.00	2029.00
13683.00	13752.00	2897.00	2881.00	3845.00	3825.00
7777.00	7675.00	736.00	714.00	3180.00	3192.00
519.00	508.00	14213.00	14297.00	1451.00	1448.00
769.00	759.00	1041.00	1082.00	5371.00	5459.00
33833.00	30657.00	5564.00	5453.00	646.00	632.00
5276.00	5201.00	1659.00	1648.00	748.00	737.00
2816.00	2771.00	26979.00	24815.00	4445.00	4419.00
7777.00	7750.00	6242.00	6209.00	2326.00	2308.00
2935.00	2913.00	1041.00	1024.00	5045.00	5023.00
2816.00	2827.00	3666.00	3630.00	8708.00	8619.00

1221.00	1210.00	3949.00	3923.00	527.00	531.00
890.00	878.00	777.00	767.00	791.00	791.00
3369.00	3317.00	1483.00	1470.00	646.00	643.00
658.00	641.00	3949.00	3851.00	3650.00	3525.00
19257.00		5656.00	5567.00	1532.00	1538.00
2071.00		930.00	917.00	4804.00	4766.00
2816.00		930.00	918.00	3650.00	3699.00
		4033.00	3969.00	5794.00	5678.00
		1041.00	1017.00	17463.00	17311.00
		8942.00		2690.00	2655.00
				1864.00	1835.00
				1532.00	1532.00
				1451.00	1437.00
				14855.00	14780.00
				6080.00	6005.00
				1532.00	1482.00
				28565.00	

Unmatched fragments are due to vector-containing sequences. We expect 3, 1 and 1-2 such fragments in the BgIII, EcoRI, and NsiI enzyme domains, respectively. The sum of these fragments can be calculated from the sequence of the BAC clone and, in this clone, should be 26,152bp, 9,439bp, and 25,563bp for the three enzyme domains.

## FEATURES

### source

Location/Qualifiers  
 1..179757  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="p14-15"  
 /clone="De Jong: dj182j23"  
 /cell\_line="Male blood"  
 /clone\_lib="De Jong BAC Library"  
 2009..2065  
 /rpt\_family="Alu"  
 3184..3272  
 /rpt\_family="MER20"  
 3891..4165  
 /rpt\_family="Alu"  
 complement(4937..5211)  
 /rpt\_family="Alu"  
 complement(6618..6908)  
 /rpt\_family="Alu"  
 complement(7575..7871)  
 /rpt\_family="Alu"  
 complement(8397..8669)  
 /rpt\_family="Alu"  
 10475..10554  
 /rpt\_family="MT1"  
 complement(13453..13492)  
 /rpt\_family="Alu"  
 14584..14861  
 /rpt\_family="Alu"  
 complement(15593..15880)  
 /rpt\_family="Alu"  
 16455..16794  
 /rpt\_family="THE1"

```

repeat_region 16799..18673
/rpl_family="MSTAR"
repeat_region /rpl_family="Alu" complement(17201..17498)
repeat_region 17925..17966
/rpl_family="THE1"
repeat_region 18711..19051
/rpl_family="THE1"
repeat_region 19606..19875
/rpl_family="Alu"
repeat_region 19885..20236
/rpl_family="THE1"
repeat_region complement(23443..23690)
/rpl_family="Alu"
repeat_region complement(23917..24133)

```

Query Match 16.2% Score 236; DB 42; Length 179757;  
 Best Local Similarity 84.5%; Pred. No. 3.3e-45;  
 Matches 277; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

```

OY 1132 AAGAGAAATCAAGACTCTACAAAGAGGCTGGGAGGCTGATGCTGATATCC 1191
DB 50074 AAAATGATCTTAAGATTAACAAGAGGCTGGTATGCTGATATCC 50015
OY 1192 CAGCACTTTGGAG-GCGAGCGAGCGAGATCACTTGAAGTTCAGACCGCT 1250
DB 50014 CAGCACTTTGGAGTCCGAGCGAGCGAGATCACTTGAAGTTCAGACCGCT 49955
OY 1251 GCGCAAAATGATGAAATCTCTCTACTTAAATCAAAATTTAGCTGATATGCTGC 1310
DB 49954 GACCAACATGATGAAATCCCGCTCTCTAAATCAAAATTTAGCTGATATGCTGC 49895
OY 1311 AGCGCCTCTATCCAGCTACTTGGAGGCTGAGCAGAGAAATTCCTGAATATGGA 1370
DB 49894 GGGCCCTCTATCTCCAGCTACTTGGAGGCTGAGCAGAGAAATTCCTGAATATGGA 49835
OY 1371 GCGCAGGTTGAATGAGTGAATCAGACCACTATCTCCAGCTGGGCAAGAGTAA 1430
DB 49834 GGTGAGGTTAGAGTGAATGAGTGAATCAGACCACTATCTCCAGCTGGGCAAGAGTAA 49775
OY 1431 GACTCTGTCTCAAAAAAAAAAAAAA 1458
DB 49774 GACTCTGTCTCAAAAAAAAAAAAAA 49747

```

# RESULT 11 AC004824

LOCUS AC004824 134578 bp DNA HTG 12-JUN-1998  
 DEFINITION Homo sapiens clone DJ020B21, WORKING DRAFT SEQUENCE, 4 unordered  
 pieces.  
 ACCESSION AC004824  
 NID 93213173  
 VERSION AC004824.1 GI:3213173  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 134578)  
 WATERSTON,R.H.  
 TITLE The sequence of Homo sapiens clone  
 Unpublished  
 2 (bases 1 to 134578)  
 WATERSTON,R.H.  
 TITLE Direct Submission  
 Submitted (12-JUN-1998) Genome Sequencing Center, Washington  
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
 MO 63108, USA

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 4 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 6892: contig of 6892 bp in length  
 \* 6893 gap of unknown length  
 \* 6910 17944: contig of 11035 bp in length  
 \* 17945 17961: gap of unknown length  
 \* 17962 34519: contig of 16558 bp in length  
 \* 34520 34536: gap of unknown length  
 \* 34537 134578: contig of 100042 bp in length.  
 Location/Qualifiers  
 source  
 1. 134578  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DJ020B21"

BASE COUNT 32479 a 34383 c 33834 g 33831 t 51 others  
 ORIGIN

Query Match 16.1% Score 235; DB 34; Length 134578;  
 Best Local Similarity 84.4%; Pred. No. 5.5e-45;  
 Matches 276; Conservative 0; Mismatches 50; Indels 1; Gaps 1;

```

OY 1134 ACAGAAATCAAGACTCTACAAAGAGGCTGGGAGGCTGATGCTGATATCCA 1193
DB 11958 ACATATAATTTAAATTAAGCTGGCAGAGGCTGGGAGGCTGATGCTGATATCCA 12017
OY 1194 GCACTTTGGAGG-GCGAGCGAGCGAGATCACTTGAAGTTCAGACCGCTGC 1252
DB 12018 GCACTTTGGAGGCGGAGGCTGGGAGGCTGATGCTGATGATGATGATGATG 12077
OY 1253 CCAAAATGTAATCCCTGCTCTACTTAAATCAAAAGTTCAGATATGCTGCAG 1312
DB 12078 CCACATGTAATCCCGCTCTCTACTTAAATCAAAATTCAGATATGCTGCAG 12137
OY 1313 GCGCCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATTCCTGAATATGAGG 1372
DB 12138 GTGCTGTATCCAGCTACTTGGAGGCTGAGCAGAGAAATTCCTGAATATGAGG 12197
OY 1373 CAGAGGTTGAATGAGTGAATCAGACCACTATCTCCAGCTGGGCAAGAGTAA 1432
DB 12198 CGGAGGTTGAGTGAATGAGTGAATCAGACCACTATCTCCAGCTGGGCAAGAGTAA 12257
OY 1433 CTCTGTCTCAAAAAAAAAAAAAA 1459
DB 12258 CTCAGTCTCAAAAAAAAAAAAAA 12284

```

## RESULT 12

LOCUS AC005674 220000 bp DNA HTG 16-JUN-1999  
 DEFINITION Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 9 unordered  
 pieces.  
 ACCESSION AC005674  
 NID 94803911  
 VERSION AC005674.7 GI:4803911  
 KEYWORDS HTG; HTGS\_PHASE1.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 220000)  
 STONE,N.E., SCHMUTZ,J.J., COX,D.R. and MYERS,R.M.  
 TITLE Direct Submission  
 Unpublished  
 2 (bases 1 to 220000)  
 STONE,N.E., SCHMUTZ,J.J., COX,D.R. and MYERS,R.M.  
 TITLE Direct Submission  
 Submitted (11-SEP-1998) Department of Genetics, Stanford Human  
 Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA  
 On May 12, 1999 this sequence version replaced gi:4757680.

## COMMENT

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 9 contigs. The true order of the pieces



```
gene
R01243: supported by GENSCAN
/evidence-not_experimental
/product="d330M3.2 (novel protein)"
5119. 9169
/gene="d330M3.2"
join(<5119. 5253,7768. 7893)
/gene="d330M3.2"
/note="supported by GENSCAN"
/codon_start=1
/evidence-not_experimental
/product="d330M3.2 (novel protein)"
/protein_id="CAA21140.1"
/db_xref="PID:e1373176"
/db_xref="PID:g4210352"
/db_xref="GI:4210352"
/translation="AKEEFHGDYKQFLHVLRSKDKGIYVNPQSVFLIDRQHL
QTPKNAITFKLCSTLCYLPQDLTHMANGTTEHDLRPMPE"
5393. 3674
/note="Alu5g repeat: matches 1. .291 of consensus"
5715. 5763
/note="MER5B repeat: matches 1. .49 of consensus"
5883. 6000
/note="MER5A repeat: matches 37. .189 of consensus"
6010. 6308
/note="AluY repeat: matches 1. .295 of consensus"
6384. 6667
/note="AluJo repeat: matches 5. .296 of consensus"
6740. 6878
/note="MIR repeat: matches 69. .207 of consensus"
7106. 7400
/note="AluJb repeat: matches 1. .297 of consensus"
8676. 8721
/note="23 copies 2 mer tt 78% conserved"
9092. 9129
/note="19 copies 2 mer aa 82% conserved"
9124. 9129
/gene="d330M3.2"
9169
/gene="d330M3.2"
9439. 9621
/note="MIR repeat: matches 8. .256 of consensus"
9691. 9798
/note="L1MC4 repeat: matches 7851. .7977 of consensus"
10020. 10149
/note="L1MD3 repeat: matches 7361. .7499 of consensus"
10507. 10716
/note="L1MC4 repeat: matches 6633. .6858 of consensus"
10855. 11115
/note="AluJo repeat: matches 1. .283 of consensus"
11469. 11619
/note="MLT1A1 repeat: matches 1. .146 of consensus"
11620. 11906
/note="AluXx repeat: matches 1. .287 of consensus"
11907. 12055
/note="MLT1A1 repeat: matches 146. .310 of consensus"
complement(join(12316. 12571,15963. >16147))
/gene="d330M3.1"
/note="mach: ESTs_A5522110 AA207689 AA819653 AA928745
A1017897 AA993303 A1216348 AA832106 AA121562 AA552456
AA629986 AA614605 T55234 AA121366 FT17737 AA445103 AA073807
AA121372 A1052396 AA939005 A1267851 T53974; supported by
GENSCAN and FGENES"
/evidence-not_experimental
/product="d330M3.1 (novel protein similar to (predicted)
plant, worm, yeast and archaea bacterial proteins)"
complement(12316)
/gene="d330M3.1"
complement(12316. 16147)
/gene="d330M3.1"
complement(12331. 12336)
/gene="d330M3.1"
complement(join(12415. 12571,15963. >16147))
/gene="d330M3.1"
polyA_site
polyA_signal
CDS
```

```
/note="match: proteins P87304 P93828 CE02501 CE00158
P95914: supported by GENSCAN and FGENES"
/codon_start=1
/evidence-not_experimental
/product="d330M3.1 (novel protein similar to (predicted)
plant, worm, yeast and archaea bacterial proteins)"
/protein_id="CAA21139.1"
/db_xref="PID:e1373175"
/db_xref="PID:g4210351"
/db_xref="GI:4210351"
/translation="ITLVSAAPGKVCMEKYEETHTNIGTLHGTLATLVNISTMA
LCTERGAQVSVDMNITVYNSPAKGEIVITAVHLKQKTLAFTSDLTRKAKGLI
AOGRTKHLGN"
13295. 13604
/note="AluXx repeat: matches 1. .310 of consensus"
14521. 14821
/note="AluJb repeat: matches 1. .299 of consensus"
15123. 15387
/note="AluXx repeat: matches 30. .298 of consensus"
15514. 15624
/note="AluJo repeat: matches 2. .136 of consensus"
16957. 17250
/note="AluY repeat: matches 1. .294 of consensus"
17526. 17589
/note="MIR repeat: matches 194. .258 of consensus"
17534. 17637
/note="L2 repeat: matches 2588. .2692 of consensus"
17943. 18245
/note="AluJas repeat: matches 1. .311 of consensus"
18551. 18684
/note="L2 repeat: matches 2357. .2495 of consensus"
18704. 18998
/note="AluXx repeat: matches 6. .300 of consensus"
19037. 19404
/note="match: GSS B40022"
19504. 19639
/note="MIR repeat: matches 88. .238 of consensus"
20030. 20303
/note="AluXx repeat: matches 24. .297 of consensus"
complement(20331. 20463)
/note="match: GSS B74972"
21207. 21337
/note="MIR repeat: matches 90. .244 of consensus"
21347. 21600
Query Match 16.1% Score 234.6 DB 9: Length 107200:
Best Local Similarity 77.7% Pred.No. 6:7e-45:
Matches 296; Conservative 0; Mismatches 84; Indels 1; Gaps 1;
QY 1080 TGTACCCAGAGGAAACAGTACACAGATTCATAGAGTGAACAGACGAAGAACAGCA 1139
DB 83884 TCTAGCTGGGTGACAGATGAGCTCTCTATATAAATAATATAATATAGGTGAC 83825
QY 1140 AAAATCAAGACTCTACAAAGAGGTGGCGAGGTGGCTCCTGTATATCCACACTT 1199
DB 83824 AAAAATTCATATAAATAATGTGGGTGAGTGGCTCACACCTGTATATCCACACTT 83765
QY 1200 TGGGAGGC-C-GAGCGAGCGATCATCTGTAGTGAAGAGTTCAGACACCGCTGGCCAAA 1258
DB 83764 TGGGAGGCTGAGGTGGCGAGTCACTTAAGTTCAGAGGCTCAATCTGGCCAAA 83705
QY 1259 TGGTGAATCTGCTGTCTGTCTATAAATAAAGTATAGTGTATAGTGGAGGCGCT 1318
DB 83704 TGGTGAATCCCTTTCTCTTTATAAATAAATAATAGCAGATATGTCATGCGCT 83645
QY 1319 GTAATCCAGCTACTTGGAGGCTGAGCGAGAGATTCCTGAATATAGGAGGAGAG 1378
DB 83644 GTAATCCAGCTACTTGGAGGCTGAGCGAGAGATTCCTGAATATAGGAGGAGAG 83585
QY 1379 TTGAATGTAGTTGAGATCAACACCATATATCTCAGCTGGGCGACAGAGTAAGACTGT 1438
DB 83584 TTGCACTGAGCTGAGATTCACACCATCTCAGTCTGGGTGAGAGAGCAAGATTCTGT 83525
```



repeat_region	/rpt_family="AIuSg"	complement(14592..14896)
repeat_region	/rpt_family="AIuSg"	15215..15420
repeat_region	/rpt_family="L1PA2"	15439..15570
repeat_region	/rpt_family="AIuSc"	15576..15964
repeat_region	/rpt_family="L1PA2"	15985..16154
repeat_region	/rpt_family="AIuSg/x"	16188..16423
repeat_region	/rpt_family="AIuSg/x"	complement(16845..17079)
repeat_region	/rpt_family="MER47A"	17137..17231
repeat_region	/rpt_family="MER4_internal"	17271..17756
repeat_region	/rpt_family="MER9"	complement(17856..17990)
repeat_region	/rpt_family="MER85"	17994..18235
repeat_region	/rpt_family="MER4_internal"	18062..18793
repeat_region	/rpt_family="MER31_internal"	18244..18961
repeat_region	/rpt_family="MER4_internal"	19480..19549
repeat_region	/rpt_family="L2"	complement(19734..19786)
repeat_region	/rpt_family="AT_rich"	19857..20203
repeat_region	/rpt_family="AIuJo"	20224..20473
repeat_region	/rpt_family="MER31_internal"	20527..20885
repeat_region	/rpt_family="MER65_internal"	20894..21195
repeat_region	/rpt_family="AIuSx"	21235..21528
repeat_region	/rpt_family="MER65_internal"	21535..22866
repeat_region	/rpt_family="L1PA13"	23008..23307
repeat_region	/rpt_family="AIuSc"	23338..23509
repeat_region	/rpt_family="MER65_internal"	23532..23975
repeat_region	/rpt_family="MUT2A"	23977..24473
repeat_region	/rpt_family="MER65_internal"	24474..24868
repeat_region	/rpt_family="MER65A"	complement(25542..25682)
repeat_region	/rpt_family="MIR"	25986..26154
repeat_region	/rpt_family="LRR22"	26180..26394
repeat_region	/rpt_family="LRR22"	complement(26739..27076)
repeat_region	/rpt_family="MER7A"	

Query Match	16.18;	Score 234.6;	DB 11;	Length 80507;
Best Local Similarity	85.88;	Pred. No. 6.6e-45;		
Matches 272;	Conservative	0;	Mismatches 44;	Indels 1;
				Gaps 1;

Oy	1143	ATCAAGACTACAAAAGGCTGGGAGGGTGGCTCATSCGTATATCCAAACATTGG	1202
Dd	4001	ATMAAATTATCCAATCTGGGGCGGCCAATGGCTCATCTTCATCAATCCACACTTTGG	4060
Oy	1203	GAGGC-AGCGACGACATCATCTTGAGGTNAGGATTCAAGACCAGCTGGCCAAATGG	1261
Dd	4061	GAGGCCAAGGTGGGCAATACCTGTGAGTATGAGGAGTTTAGACCAAGCTGGGCCAACATGG	4120

QY	1262	TGATTCCTCTGTACTATAAATAAAGTACTGTGATATGTGTGACAGGCCCTCTGA	1321
Db	4121	CGAAACCTCTCTCTATATAAATAATACAAAAATTAAGTGGGCTGGTGGCAATGCCCTCTGA	4180
QY	1322	ATCCCACTCTCTGGAGGCGCTAGCGACAGAGAATTCCTGTAATATGGAGGACAGAGTTG	1381
Db	4181	ATCCCACTCTCTGGAGGCGCTAGCGACAGAGAATTCCTGTAATATGGAGGACAGAGTTG	4240
QY	1382	AAGTGAAGTATGATATCACCACCTACTACTCTGAGCTGGGGCAACAGATGAGACTCTGTCTC	1441
Db	4241	TAGTGAAGTATGATATGAGCACTGACCTGAGCTGGGGCAACAGATGAGACTCTGTCTC	4300
QY	1442	AAAAAAAAAAAAAAAAAAAA 1458	
Db	4301	AAAAAAAAAAAAAAAAAAAA 4317	
RESULT	15		
AC004971			
LOCUS			
DEFINITION	AC004971. 188477 bp DNA	HTG	12-JUN-1998
ACCESSION	AC004971		
MID	93213026		
VERSION	AC004971.1	GI:3213026	
KEYWORDS	HTG: HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 188477)		
REFERENCE	The sequence of Homo sapiens clone		
JOURNAL	Unpublished		
AUTHORS	2 (bases 1 to 188477)		
TITLE	Waterston,R.H.		
JOURNAL	Submitted (12-JUN-1998) Genome Sequencing Center, Washington		
COMMENT	University School of Medicine, 4444 Forest Park Parkway, St. Louis		
	MO 63108, USA		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 21 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	1	1820: contig of 1820 bp in length	
	*	1821	1839: gap of unknown length
	*	1840	3311: contig of 1472 bp in length
	*	3312	3350: gap of unknown length
	*	3331	5118: contig of 1788 bp in length
	*	5138	5137: gap of unknown length
	*	5139	6628: contig of 1491 bp in length
	*	6629	6647: gap of unknown length
	*	6648	8860: contig of 2213 bp in length
	*	8861	8879: gap of unknown length
	*	8880	10635: contig of 1756 bp in length
	*	10636	10654: gap of unknown length
	*	10655	13630: contig of 2976 bp in length
	*	13631	13649: gap of unknown length
	*	13650	17252: contig of 3603 bp in length
	*	17253	17271: gap of unknown length
	*	17272	20154: contig of 2883 bp in length
	*	20155	20173: gap of unknown length
	*	20174	24721: contig of 4548 bp in length
	*	24722	24740: gap of unknown length
	*	24741	30872: contig of 6132 bp in length
	*	30873	30881: gap of unknown length
	*	30882	36985: contig of 6094 bp in length
	*	36986	37004: gap of unknown length

FEATURES  
source  
1. 188477  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="D01135K23"  
/cdate="D01135K23"  
Location/Qualifiers  
188477: contig of 1619 bp in length.  
186859  
185141  
185123  
185122  
128056  
128055  
105584  
84183  
84165  
67906  
67887  
50134  
50133  
40431  
40412  
37005

BASE COUNT 43810 a 48919 c 48642 g 46728 t 378 others  
ORIGIN

Query Match 16.18; Score 234.6; DB 34; Length 188477;  
Best Local Similarity 84.38; Pred. No. 7.1e-45;  
Matches 264; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 1146 AAGACTTACAAAGAGCGCTGGCGAGGCTGCTCATGCTGTAATCCAGACATTGGGAG 1205  
DB 86621 AGGATATTGCAAAAGCGCTGGCGAGGCTGCTCATGCTGTAATCCAGACATTGGGAG 86680  
OY 1206 GCGAGCGAGCGAGTCACTTGAAGAGTTCAGAGCCAGCTGGCCAAATGTGAA 1265  
DB 86681 CCAAGCGCGAGTCACTTGAAGAGTTCAGAGCCAGCTGGCCAAATGTGAA 86740  
OY 1266 ATCTGCTCTACTAAATAATCAAAAGTGTGATATGTTGGCAGCGCTGTAATCC 1325  
DB 86741 ACCCTGCTCTACTAAATAATCAAAAGTGTGATATGTTGGCAGCGCTGTAATCC 86800  
OY 1326 CAGCTACTTGGAGGCTGAGGAGAGATGCTTGAATATGGAGGAGAGGTTGAAGT 1385  
DB 86801 CAGCTACTTGGAGGCTGAGGAGAGATGCTTGAATATGGAGGAGAGGTTGAAGT 86860  
OY 1386 GAGTTGAGATCACACCTACTCTCAGCTGGGCAACAGAGTAAAGTCTGTCTCAAA 1445  
DB 86861 GAGCGGAGATGCCCGCACTGCACTCAGCGGCAAGAGTGAAGTCTGTCTCAAA 86920  
OY 1446 AAAAAAAAAAAAAA 1458  
DB 86921 AAAAAAAAAAAAAA 86933

Search completed: September 28, 1999, 12:34:03  
Job time: 5723 sec

**This Page Blank (uspto)**





```
|||||
Db 421 NGAGGCTCAGTAACTGATGACCCGCTGTACACCCCAACATGTTGCGCCGCG 480
Oy 481 AGGGCAAGACAGAGAGTCTCTGCAACGTGAGAGAGGGGAAAAGGGAGGCGGACT 540
Db 481 AGGGCAAGACAGAGAGTCTCTGCAACGTGAGAGAGGGGAAAAGGGAGGCGGACT 540
Oy 541 CAGGGAAGGTGAGAGAGGGGAGACAGACACAGGGGCCCATGGCGAGATGACAG 600
Db 541 CAGGGAAGGTGAGAGAGGGGAGACAGACACAGGGGCCCATGGCGAGATGACAG 600
Oy 601 ATGAGAGACACACAGAGAGACAGTACCACTAGAGAGAGAACTAGAGAAACAGAGA 660
Db 601 ATGAGAGACACACAGAGAGACAGTACCACTAGAGAGAGAACTAGAGAAACAGAGA 660
Oy 661 ATAAACACAGAGATTAAGAGAGCAAAAGAGAGAGAAACAGAAACAGACATGGAGGC 720
Db 661 ATAAACACAGAGATTAAGAGAGCAAAAGAGAGAGAAACAGAAACAGACATGGAGGC 720
Oy 721 AGAACAACACACATAGAAATGCACTTGCCTCCACACAGCATGGGGCCTGAGGCGGT 780
Db 721 AGAACAACACACATAGAAATGCACTTGCCTCCACACAGCATGGGGCCTGAGGCGGT 780
Oy 781 GACCTCCACCCCAATAGAAAATCCTCTTATTAACCTTTGACTCCCAAAAACCTGACTAGA 840
Db 781 GACCTCCACCCCAATAGAAAATCCTCTTATTAACCTTTGACTCCCAAAAACCTGACTAGA 840
Oy 841 ATAGCTTCTGTGAGAGGGGAGGCTTACCAATTAACATTAATGCTATTTATCATACGT 900
Db 841 ATAGCTTCTGTGAGAGGGGAGGCTTACCAATTAACATTAATGCTATTTATCATACGT 900
Oy 901 TTATGCTATCATATATACCTTTGTTGAAATTTTGTGATTTTCAAGTACACAGTTC 960
Db 901 TTATGCTATCATATATACCTTTGTTGAAATTTTGTGATTTTCAAGTACACAGTTC 960
Oy 961 GTCTGTGAATTTTAAATTTGTTGCAACTCTCTTAAATTTTCTGATGTGTTATTGA 1020
Db 961 GTCTGTGAATTTTAAATTTGTTGCAACTCTCTTAAATTTTCTGATGTGTTATTGA 1020
Oy 1021 AAAAATCCAGTAAAGTGTGACTTTGTCATTAACCAAGGGTGTTCAGAGGCTCAACTGT 1080
Db 1021 AAAAATCCAGTAAAGTGTGACTTTGTCATTAACCAAGGGTGTTCAGAGGCTCAACTGT 1080
Oy 1081 GTACCCAGAGAGAAACAGTACAGACAGATTCATAGAGGTGAACAGAGAAAGAAAGGAA 1140
Db 1081 GTACCCAGAGAGAAACAGTACAGACAGATTCATAGAGGTGAACAGAGAAAGAAAGGAA 1140
Oy 1141 AATCAAGACTTACAAAGAGGCTGGGAGGCTGCTCATGCTGTATTCAGCACTTT 1200
Db 1141 AATCAAGACTTACAAAGAGGCTGGGAGGCTGCTCATGCTGTATTCAGCACTTT 1200
Oy 1201 GGGAGGCGAGGCGAGATCACTTGAAGTAAAGAGTTCAAGACAGGCTGGCCAAATG 1260
Db 1201 GGGAGGCGAGGCGAGATCACTTGAAGTAAAGAGTTCAAGACAGGCTGGCCAAATG 1260
Oy 1261 GTGAATTCCTGTCTACTTAAATACAAAGTTAGCTGATATGTTGGGAGGCGCTGT 1320
Db 1261 GTGAATTCCTGTCTACTTAAATACAAAGTTAGCTGATATGTTGGGAGGCGCTGT 1320
Oy 1321 AATCCAGCTACTTGGAGGCTGAGGCGAGAGAAATGCTTGAATATGAGAGCAGAGTT 1380
Db 1321 AATCCAGCTACTTGGAGGCTGAGGCGAGAGAAATGCTTGAATATGAGAGCAGAGTT 1380
Oy 1381 GAAGTGAAGTGAATCAACACACTATATCTCAGCTGGGGCAACAGAGTAAAGACTGTCT 1440
Db 1381 GAAGTGAAGTGAATCAACACACTATATCTCAGCTGGGGCAACAGAGTAAAGACTGTCT 1440
Oy 1441 CAAAAAAGAAAAAAGAAAA 1459
Db 1441 CAAAAAAGAAAAAAGAAAA 1459
```

RESULT: 2.

```
V61251
ID V61251 standard; cDNA; 1459 bp.
AC V61251.
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; 003492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PA 01-AUG-1997; US-904804.
PI (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 106; 130p; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SO Sequence 1459 bp; 427 A; 328 C; 406 G; 295 T;

Query Match 99.8%; Score 1456; DB 1; Length 1459;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GGTACGCGCACAGTGTTCAGAGTGAAGTGCAGAGCTCTACACCAATCGGGCTGGCC 60
Db 1 GGTACGCGCACAGTGTTCAGAGTGAAGTGCAGAGCTCTACACCAATCGGGCTGGCC 60
Oy 61 TGCACAGTCTTGAAGCCGACCAAGAGCGAGGAGCCAGATGTGAGAGCCAGCTCTCG 120
Db 61 TGCACAGTCTTGAAGCCGACCAAGAGCGAGGAGCCAGATGTGAGAGCCAGCTCTCG 120
Oy 121 TACGGCACCCAGAGTACCAACAGACCTTGTCTGTACGACCTCATGCTCAATGAATTGG 180
Db 121 TACGGCACCCAGAGTACCAACAGACCTTGTCTGTACGACCTCATGCTCAATGAATTGG 180
Oy 181 AGGAATCCGTGTCCAGTCTGACACCATCCGAGAGATAGCATTTCTTCCAGTCCCTA 240
Db 181 AGGAATCCGTGTCCAGTCTGACACCATCCGAGAGATAGCATTTCTTCCAGTCCCTA 240
Oy 241 CCGGGGGAACCTTGTCCGCTTGTGCTGGGCTGTGCTGCGAAGCGTGAAGTCAACGG 300
Db 241 CCGGGGGAACCTTGTCCGCTTGTGCTGGGCTGTGCTGCGAAGCGTGAAGTCAACGG 300
Oy 301 GTGTGTGTCTGCCCTTTCAGAGAGTCTCTGCCAGTCCGGGGGCTGACCCAGAGCT 360
Db 301 GTGTGTGTCTGCCCTTTCAGAGAGTCTCTGCCAGTCCGGGGGCTGACCCAGAGCT 360
Oy 361 CTGCGTCCAGGCAAGTCCATACCGTCTGTGAGTGTGTAACGTTGTGGTGTCTGA 420
Db 361 CTGCGTCCAGGCAAGTCCATACCGTCTGTGAGTGTGTAACGTTGTGGTGTCTGA 420
Oy 421 NGAGGCTCAGTAACTATAGACCCGCTGTACACCCCAACATGTTCTGGCGGGGG 480
Db 421 NGAGGCTCAGTAACTATAGACCCGCTGTACACCCCAACATGTTCTGGCGGGGG 480
Oy 481 AGGGCAAGACAGAAAGACTCTCTGCAACGTGAGAGAGGGGAAAAGGGAGGCGAGCT 540
Db 481 AGGGCAAGACAGAAAGACTCTCTGCAACGTGAGAGAGGGGAAAAGGGAGGCGAGCT 540
Oy 541 CAGGGAAGGTGAGAGAGGGGAGACAGAGACACAGGGCCGATGGCGAGATGCAAG 600
Db 541 CAGGGAAGGTGAGAGAGGGGAGACAGAGACACAGGGCCGATGGCGAGATGCAAG 600
Oy 601 ATGAGAGACACACAGAGAGACAGTACCACTAGAGAGAGAACTAGAGAAACAGAGA 660
```

```

|||||
Db 601 ATGAGAGACACACAGGAGACACTGACACTAGAGAGAACTGAGAGAAACAGAGAA 660
Oy 661 ATAAACACAGGATTAAGAGAGAAAGCAAGAGAGAAACAGAAACATGGGGAGCC 720
Db 661 ATAAACACAGGATTAAGAGAGAAAGCAAGAGAGAAACAGAAACATGGGGAGCC 720
Oy 721 AGAAACACACACACATAGAGAAAGTGGAGCTTCCACACAGCATGGGGCTGAGGGCGGT 780
Db 721 AGAAACACACACACATAGAGAAAGTGGAGCTTCCACACAGCATGGGGCTGAGGGCGGT 780
Oy 781 GACCTCCACCAATAGAAATCTCTTATTAATCTTGAAGCTCCCAAAAACCTGACTAGAA 840
Db 781 GACCTCCACCAATAGAAATCTCTTATTAATCTTGAAGCTCCCAAAAACCTGACTAGAA 840
Oy 841 ATAGCTACTGTGAGGGGGAGCTTCCATTAATAGTGAATGATTTGATAGCT 900
Db 841 ATAGCTACTGTGAGGGGGAGCTTCCATTAATAGTGAATGATTTGATAGCT 900
Oy 901 TTATGATCATGATATACCTTTGTTGGAAATTTTGTATTTCTAAGTACAGATTC 960
Db 901 TTATGATCATGATATATACCTTTGTTGGAAATTTTGTATTTCTAAGTACAGATTC 960
Oy 961 GTCGTGAAATTTTAAATGTTGCAACTCTCTAAATTTTCTGATGTTATTGA 1020
Db 961 GTCGTGAAATTTTAAATGTTGCAACTCTCTAAATTTTCTGATGTTATTGA 1020
Oy 1021 AAAATCCAAATATAGTGGAGCTTGTGATCAAAACCGGTTGTTCAAGGTTCACTGT 1080
Db 1021 AAAATCCAAATATAGTGGAGCTTGTGATCAAAACCGGTTGTTCAAGGTTCACTGT 1080
Oy 1081 GTACCCAGAGGAGAAACAGTACAGAGATTCATAGAGTGAAGCAAGAGAGAGAGAA 1140
Db 1081 GTACCCAGAGGAGAAACAGTACAGAGATTCATAGAGTGAAGCAAGAGAGAGAGAA 1140
Oy 1141 AAATCAAGACTCTCAAAAGAGGCTGGGCGAGGCTGCTCATCTGTAATCCACACTTT 1200
Db 1141 AAATCAAGACTCTCAAAAGAGGCTGGGCGAGGCTGCTCATCTGTAATCCACACTTT 1200
Oy 1201 GGGAGGGAGGAGGAGGAGTACTGAGTGAAGAGATTCAGAGACCAAGCTGGCCAAATG 1260
Db 1201 GGGAGGGAGGAGGAGGAGTACTGAGTGAAGAGATTCAGAGACCAAGCTGGCCAAATG 1260
Oy 1261 GTGAATCTCTGTACTTAAATAAATCAAAAGTATAGTGTGGAGGGGCGCTGT 1320
Db 1261 GTGAATCTCTGTACTTAAATAAATCAAAAGTATAGTGTGGAGGGGCGCTGT 1320
Oy 1321 AATCCAGCTACTTGGAGGCTGAGGAGGAGAAATGCTGTAATATGGAGGAGAGTT 1380
Db 1321 AATCCAGCTACTTGGAGGCTGAGGAGGAGAAATGCTGTAATATGGAGGAGAGTT 1380
Oy 1381 GAATGAGTTAGATACACACACTATCTCAAGCTGGGGCAACAGAGTAACTGTCT 1440
Db 1381 GAATGAGTTAGATACACACACTATCTCAAGCTGGGGCAACAGAGTAACTGTCT 1440
Oy 1441 CAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1459
Db 1441 CAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1459

```

## RESULT 3

V58645  
ID V58645 standard: cDNA: 1265 BP.  
AC V58645: 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE2.  
KM Prostate tumour specific gene: human; prostate cancer; detection;  
OS Homo sapiens.  
PN W09837418-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998: U03690.  
PR 09-FEB-1998: US-904809.

```

PR 25-FEB-1997: US-806596.
PR 01-AUG-1997: US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC. Xu J.
DR WPI: 98-480805/41.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1: Page 113-114, 141pp: English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1265 BP: 256 A: 432 C: 321 G: 245 T:

```

Query Match 32.5%; Score 474; DB 1; Length 1265;  
Best Local Similarity 98.4%; Pred. NO. 4.4e-95;  
Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

Oy 25 AGTGAAGTGCAGAGCTCTACACACATCGGGCTGGGCTGCACAGCTTGAGGCGACCAAG 84
Db 104 ACTGTTCCAGAACTCTACACATCGGGCTGGGCTGCACAGCTTGAGGCGACCAAG 163
Oy 85 AGCCAGGAGCCAGAGTGTGAGGCGCAGCTCTCCGTAAGGCGACCAAGTACAAAGAC 144
Db 164 AGCCAGGAGCCAGAGTGTGAGGCGCAGCTCTCCGTAAGGCGACCAAGTACAAAGAC 223
Oy 145 CCTGTCTGCTAACAGACTCATGCTCATCAAGTGAAGCAATCCGTGTCGAGTGTGACA 204
Db 224 CCTGTCTGCTAACAGACTCATGCTCATCAAGTGAAGCAATCCGTGTCGAGTGTGACA 283
Oy 205 CCATCCGAGAGATGAGCATTCGTCGAGTGCCTTACCGGGGAGAACTTGGCTCGTTT 264
Db 284 CCATCCGAGAGATGAGCATTCGTCGAGTGCCTTACCGGGGAGAACTTGGCTCGTTT 343
Oy 265 CTGGCTGGGCTCTGTCGCGAAGCGGTAGCTACAGGCTGTGTGTCCTTTCAAGGA 324
Db 344 CTGGCTGGGCTCTGTCGCGAAGCGGTAGCTACAGGCTGTGTGTCCTTTCAAGGA 403
Oy 325 GGTCTCTGCCAATCGCGGGGGGTGACCAAGACTTGGTCCAGGCAAGATGCTTAC 384
Db 404 GGTCTCTGCCAATCGCGGGGGGTGACCAAGACTTGGTCCAGGCAAGATGCTTAC 463
Oy 385 CGTGTGAGTGCAGTGAAGCTGTGCTGTGTGAGAGAGTGTGANTAAAGCTCTATGA 444
Db 464 CGTGTGAGTGCAGTGAAGCTGTGCTGTGTGAGAGAGTGTGANTAAAGCTCTATGA 523
Oy 445 CCCGCTGTACACCCCAACATGTTCTGCGCGGGGAGGCGCAAGACAGAAAGACTCTG 504
Db 524 CCCGCTGTACACCCCAACATGTTCTGCGCGGGGAGGCGCAAGACAGAAAGACTCTG 583
Oy 505 CAACG 509
Db 584 CAACG 588

```

## RESULT 4

V61250  
ID V61250 standard: cDNA: 1265 BP.  
AC V61250:  
DT 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DE2.  
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN W09837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998: U03492.  
PR 09-FEB-1998: US-020956.

PR 25-FEB-1997; US-806099.  
 PR 01-AUG-1997; US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR MPI: 98-609886/51.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 used in a vaccine for the treatment of prostate cancer  
 PS Claim 3; Page 105-106; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 can be used as a vaccine for the treatment of prostate cancer. The DNA  
 was identified by analysis of a subcloned cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 32.5%; Score 474; DB 1; Length 1265;  
 Best Local Similarity 98.4%; Pred. No. 4.4e-95;  
 Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 25 AGTGAGTGCAGAGCTCTTACACCATGGGGCTGGCAGACTCTTGAGCGCGACCAAG 84  
 DB 104 ACTGTTCTCGAAGCTCTTACACCATGGGGCTGGCAGACTCTTGAGCGCGACCAAG 163  
 QY 85 AGCCAGGAGACCATAGTGTGAGAGCCAGCTCTCCGACGAGCAGCAGATACAGAGAC 144  
 DB 164 AGCCAGGAGACCATAGTGTGAGAGCCAGCTCTCCGACGAGCAGCAGATACAGAGAC 223  
 QY 145 CCTTGTCTGCTAAGACCTCATCTCATCAAGTTGAGAGATCCGTGTCGAGTGTGACA 204  
 DB 224 CCTTGTCTGCTAAGACCTCATCTCATCAAGTTGAGAGATCCGTGTCGAGTGTGACA 283  
 QY 205 CCAATCCGGAGACATGAGATGCTTCCGAGTCCCTACCGGGGGAGACTCTTGGCTGCTT 264  
 DB 284 CCAATCCGGAGACATGAGATGCTTCCGAGTCCCTACCGGGGGAGACTCTTGGCTGCTT 343  
 QY 265 CTGCTGTGGGCTGTGCTGCGACAGCGTGAAGTCAAGGGTGTGTCTGCTTCAAGGA 324  
 DB 344 CTGCTGTGGGCTGTGCTGCGACAGCGTGAAGTCAAGGGTGTGTCTGCTTCAAGGA 403  
 QY 325 GGTCTCTGCTGCTGCTGCGACAGCGTGAAGTCAAGGGTGTGTCTGCTTCAAGGA 384  
 DB 404 GGTCTCTGCTGCTGCTGCGACAGCGTGAAGTCAAGGGTGTGTCTGCTTCAAGGA 463  
 QY 385 CGTGTGCAAGTGTGCTGCTGCGAGTGTGAGAGTGTGAGATGAGTCTATGA 444  
 DB 464 CGTGTGCAAGTGTGCTGCTGCGAGTGTGAGAGTGTGAGATGAGTCTATGA 523  
 QY 445 CCGGCTGTACACCCGACATGTTCTGCGCGGAGAGGAGGACAGAGAGTCTCTG 504  
 DB 524 CCGGCTGTACACCCGACATGTTCTGCGCGGAGAGGAGGACAGAGAGTCTCTG 583  
 QY 505 CAACG 509  
 DB 584 CAACG 588

RESULT 5  
 V11855  
 ID V11855 standard; cDNA: 1386 BP.  
 AC V11855;  
 DT 11-SEP-1998 (first entry)  
 DE Homo sapiens Tub Interactor (htr-1) gene.  
 KW serine protease; tub interactor; treatment; obesity; cachexia;  
 KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;  
 KW neurodegenerative disease; Alzheimer's disease; drug screening;  
 KW Parkinson's disease; Huntington's chorea; detection; diagnosis;  
 KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 2..701  
 FT CDS /tag- a

FT /product= htr-1 protein  
 FT /note= "putative serine protease"

PN W09812302-A1.  
 PD 26-MAR-1998.  
 PF 05-SEP-1997; U15627.  
 PR 21-JUL-1997; US-897340.  
 PR 17-SEP-1996; US-715032.  
 PA (MILL-) MILENNIUM PHARM INC.  
 PI Estrada PR, Gimeno CJ;  
 DR MPI: 98-217246/19.  
 DR P-PSDB: W59129.  
 PT Tub interactor genes - used to develop products for the treatment  
 PT of obesity, cachexia, anorexia nervosa or related disorders e.g.  
 PS diabetes  
 PS Claim 10; Fig 1; 120pp; English.  
 CC The sequence is that of the Tub Interactor gene htr-1 which  
 CC codes for a putative serine protease. It genes function  
 CC in biochemical pathways involved in weight control and  
 CC related disorders. The products can be used for treating  
 CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,  
 CC or a related disorder such as diabetes. The products can  
 CC also be used to modulate cell cycle progression and apoptosis.  
 CC They can be used for treating neurodegenerative diseases  
 CC which are characterised by apoptosis, including Alzheimer's  
 CC disease, Parkinson's disease, Huntington's chorea, amyotrophic  
 CC lateral sclerosis or spinocerebellar degenerations. The  
 CC products can also be used for detection, diagnosis and  
 CC drug screening.  
 SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 22.5%; Score 329; DB 1; Length 1386;  
 Best Local Similarity 83.1%; Pred. No. 2.3e-63;  
 Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTGACCCGACACATGTTTCCAGAGTGAAGTGCAGAGCTCTTACACCATCGGGCTGGCT 61  
 DB 100 GTGACCCGACACATGTTTCCAGAGTGAAGTGCAGAGCTCTTACACCATCGGGCTGGCT 159  
 QY 62 GCACAGTCTTGAGGCGGACCAAGAGCCAGAGATGTTGAGAGCGACCTTCCGT 121  
 DB 160 GCACAGTCTTGAGGCGGACCAAGAGCCAGAGATGTTGAGAGCGACCTTCCGT 219  
 QY 122 AGGACCCAGAGTCAACAGACCTTCTGCTTACAGACCTCATGCTCATCAAGTTGA 181  
 DB 220 AGGACCCAGAGTCAACAGACCTTCTGCTTACAGACCTCATGCTCATCAAGTTGA 279  
 QY 182 CGAATCCGTTCCGAGTGTGACACCATCCGAGCATCGATGCTTGGCAGTGCCTAC 241  
 DB 280 CGAATCCGTTCCGAGTGTGACACCATCCGAGCATCGATGCTTGGCAGTGCCTAC 339  
 QY 242 CCGGGGGAATCTTCCCTGCTTGTGCTGGGCTGCTGCGAAGGTTGAGTCAAGCG 301  
 DB 340 CCGGGGGAATCTTCCCTGCTTGTGCTGGGCTGCTGCGAAGGTTGAGTCAAGCG 386  
 QY 302 TGTGTGTGCTCCCTCTCAAGAGAGTCTTGCACAGTCCGGGGGCTGACCAAGACTC 361  
 DB 386  
 QY 362 TGCCTCCAGAGAGATGCTTACCTGCTGAGTGCAGTGAACGTGCTGGTGTCTGAN 421  
 DB 386 GGCAGATGCTTACCTGCTGAGTGCAGTGAACGTGCTGGTGTCTGAN 436  
 QY 422 GAGTGTGATTAAGTCTATATACCCGCTGTACACCCGACATGTTTGGCGCGCGGA 481  
 DB 437 GAGTGTGATTAAGTCTATATACCCGCTGTACACCCGACATGTTTGGCGCGCGGA 496  
 QY 482 GGGCAAGACCAAGAGACTCTGCAAGC 509  
 DB 497 GGGCAAGACCAAGAGACTCTGCAAGC 524

RESULT 6

V58644  
ID V58644 standard; cDNA; 1248 BP.  
AC V58644;  
DE 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DEL.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 217..696  
FT /tag= a  
FT MO9837418-A2.  
PN 27-AUG-1998.  
PE 25-FEB-1998; U03690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI: 98-480805/41.  
DR P-PSDB: W69387.  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
PS Claim 1: Page 112; 141pp; English.  
CC This sequence represents a human prostate tumour specific gene, and can  
CC be used in the method of the invention. The method is for detecting  
CC prostate cancer comprising contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC encoded by this sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 22.5%; Score 329; DB 1; Length 1248;  
Best Local Similarity 83.1%; Pred. No. 2.2e-63;  
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTGACGGCGCACACTGTTCCAGAGTGCACAGCTCCTACACCATCGGGCTGGGCT 61  
DB 120 GTGACGGCGCACACTGTTCCAGAGTGCACAGCTCCTACACCATCGGGCTGGGCT 179  
QY 62 GCACAGCTTGAGGCGCCAGAGAGCCAGAGGAGCCAGATGGTGAAGGCCACTCTCCGT 121  
DB 180 GCACAGCTTGAGGCGCCAGAGAGCCAGAGGAGCCAGATGGTGAAGGCCACTCTCCGT 239  
QY 122 ACGGACCCAGAGTACAAGAGACCCCTTGCTGCTACAGACCTCATGCTCATCAAGTTGA 181  
DB 240 ACGGACCCAGAGTACAAGAGACCCCTTGCTGCTACAGACCTCATGCTCATCAAGTTGA 299  
QY 182 CGAATCCGTGTCGAGTGTACACCATCCGAGATCATGCAATTCCTTGGAGTGCCTAC 241  
DB 300 CGAATCCGTGTCGAGTGTACACCATCCGAGATCATGCAATTCCTTGGAGTGCCTAC 359  
QY 242 CGGCGGGAACCTCTGCTGCTTCTGCTGGGTCTGCTGGCGAAGCGTGAAGTCAAGGG 301  
DB 360 CGGCGGGAACCTCTGCTGCTTCTGCTGGGTCTGCTGGCGAAGCGTGAAGTCAAGGG 406  
QY 302 TGTGTGTGCTGCTCTTCAAGAGAGTCTCTGCCAGTCCGGGGGCTGACCCAGAGCTC 361  
DB 406 ----- 406  
QY 362 TGCCTCCAGGACAGATGCTACCGTGTCTGAGTGCCTGAAGCTGTGGTGTCTGAN 421  
DB 406 -----GGCAGAAATGCTTACCGTGTCTGAGTGCCTGAAGCTGTGGTGTCTGAN 456  
QY 422 GAGGTGCAATTAAGCTCTATGAGCCGCTGTACACCCAGCATGTTCTGGCGGGGGA 481  
DB 457 GAGGTGCAATTAAGCTCTATGAGCCGCTGTACACCCAGCATGTTCTGGCGGGGGA 516  
QY 482 GGGCAAGACAGAAAGACTCTGCAAGC 509

DB 517 GGGCAAGACAGAAAGACTCTGCAAGC 544

RESULT 7  
ID V61249  
AC V61249;  
DE 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DEL.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN MO9837093-A2.  
PN 27-AUG-1998.  
PE 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI: 98-609886/51.  
DR P-PSDB: W71871.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
PS Claim 3: Page 104; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 22.5%; Score 329; DB 1; Length 1248;  
Best Local Similarity 83.1%; Pred. No. 2.2e-63;  
Matches 422; Conservative 0; Mismatches 3; Indels 83; Gaps 1;

QY 2 GTGACGGCGCACACTGTTCCAGAGTGCACAGCTCCTACACCATCGGGCTGGGCT 61  
DB 120 GTGACGGCGCACACTGTTCCAGAGTGCACAGCTCCTACACCATCGGGCTGGGCT 179  
QY 62 GCACAGCTTGAGGCGCCAGAGAGCCAGAGGAGCCAGATGGTGAAGGCCACTCTCCGT 121  
DB 180 GCACAGCTTGAGGCGCCAGAGAGCCAGAGGAGCCAGATGGTGAAGGCCACTCTCCGT 239  
QY 122 ACGGACCCAGAGTACAAGAGACCCCTTGCTGCTACAGACCTCATGCTCATCAAGTTGA 181  
DB 240 ACGGACCCAGAGTACAAGAGACCCCTTGCTGCTACAGACCTCATGCTCATCAAGTTGA 299  
QY 182 CGAATCCGTGTCGAGTGTACACCATCCGAGATCATGCAATTCCTTGGAGTGCCTAC 241  
DB 300 CGAATCCGTGTCGAGTGTACACCATCCGAGATCATGCAATTCCTTGGAGTGCCTAC 359  
QY 242 CGGCGGGAACCTCTGCTGCTTCTGCTGGGTCTGCTGGCGAAGCGTGAAGTCAAGGG 301  
DB 360 CGGCGGGAACCTCTGCTGCTTCTGCTGGGTCTGCTGGCGAAGCGTGAAGTCAAGGG 406  
QY 302 TGTGTGTGCTGCTCTTCAAGAGAGTCTCTGCCAGTCCGGGGGCTGACCCAGAGCTC 361  
DB 406 ----- 406  
QY 362 TGCCTCCAGGACAGATGCTACCGTGTCTGAGTGCCTGAAGCTGTGGTGTCTGAN 421  
DB 406 -----GGCAGAAATGCTTACCGTGTCTGAGTGCCTGAAGCTGTGGTGTCTGAN 456  
QY 422 GAGGTGCAATTAAGCTCTATGAGCCGCTGTACACCCAGCATGTTCTGGCGGGGGA 481  
DB 457 GAGGTGCAATTAAGCTCTATGAGCCGCTGTACACCCAGCATGTTCTGGCGGGGGA 516  
QY 482 GGGCAAGACAGAAAGACTCTGCAAGC 509  
DB 517 GGGCAAGACAGAAAGACTCTGCAAGC 544

```
RESULT 8
ID V58647 standard; cDNA: 1167 BP.
AC V58647;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13.
KM Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 28..645
FT CD5 /*tag= a
PN MO9837418-AZ.
PD 27-AUG-1998.
PF 25-FEB-1998: U03690.
PR 09-FEB-1998: US-904809.
PR 25-FEB-1997: US-806596.
PR 01-AUG-1997: US-904809.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-480805/41.
DR P-PSDB: W69388.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1: Page 115: 141pp: English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1167 BP: 242 A: 400 C: 287 G: 222 T;
```

```
DB 429 CCCGCTGTACCAACCCAGCATGTCTGCGCGCGGAGGCAAGACAGAGACTCTG 488
QY 505 CAACG 509
DB 489 CAACG 493

RESULT 9
ID V61252 standard; cDNA: 1167 BP.
AC V61252;
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.
KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
PN MO9837093-AZ.
PD 27-AUG-1998.
PF 25-FEB-1998: U03492.
PF 09-FEB-1998: US-020956.
PR 25-FEB-1997: US-806099.
PR 01-AUG-1997: US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J;
DR WPI: 98-609886/51.
DR P-PSDB: W71782.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3: Page 107: 130pp: English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1167 BP: 242 A: 400 C: 287 G: 222 T;
```

```
Query Match 20.1%; Score 292.8; DB 1; Length 1167;
Best Local Similarity 80.4%; Pred. No. 1.8e-55;
Matches 390; Conservative 0; Mismatches 12; Indels 83; Gaps 1;

QY 25 AGTGAAGTCAGAGCTCTTACACCATCGGGCTGCGACAGTCTTGAGGCCGACCAAG 84
DB 92 ACTGTTCCAGAACTCTTACACCATCGGGCTGCGACAGTCTTGAGGCCGACCAAG 151
QY 85 AGCCAGGAGCCAGATGTGGAGGCCAGCTCTCCGTACGGACCCAGAGTACACAGAC 144
DB 152 AGCCAGGAGCCAGATGTGGAGGCCAGCTCTCCGTACGGACCCAGAGTACACAGAC 211
QY 145 CCTGCTCGCTAAGCACTATGCTCATCAAGTGGAGCAATCCGTCCGAGTGTACA 204
DB 212 TCTTGCTCGCTAAGCACTATGCTCATCAAGTGGAGCAATCCGTCCGAGTGTACA 271
QY 205 CCATCCGAGACATCAGATTGCTTGCAGTGCCTTACCGGGGGAATCTTGCTCTTT 264
DB 272 CCATCCGAGACATCAGATTGCTTGCAGTGCCTTACCGGGGGAATCTTGCTCTTT 331
QY 265 CTGGCTGGGGTCTGCTGGGGAAGGTGAGCTCACGGGTGTGTGCTCCCTTTCAAGA 324
DB 332 CTGGCTGGGGTCTGCTGGGGAAC----- 355
QY 325 GGTCTCTGCCAGTGGCGGGGGCTGACCCAGAGCTCTGCTCCAGGACAGATGCTTAC 384
DB 355 -----GGCAGATGCTTAC 368
QY 385 CGTGTGCAAGTGGTGAACGTGTGCTGTGTGANGAGGTCTGCANTAACTCTATGA 444
DB 369 CGTGTGCAAGTGGTGAACGTGTGCTGTGTGANGAGGTCTGCANTAACTCTATGA 428
QY 445 CCCGCTGTACCAACCCAGCATGTCTGCGCGCGGAGGCAAGACAGAGACTCTG 504
DB 429 CCCGCTGTACCAACCCAGCATGTCTGCGCGCGGAGGCAAGACAGAGACTCTG 488
```

```

OY 505 CACG 509
    |||||
Db 489 CACG 493

RESULT 10
V37495
ID V37495 standard: DNA: 871 BP.
AC V37495:
DE 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW benign prostate hyperplasia; diagnosis; drug screening; PSK; ss.
OS Homo sapiens.
FH Key 31.777
FT CDS location/Qualifiers
    /*tag= a
    /transl_except= (pos:367..369, aa:Xaa)
    /transl_except= (pos:412..414, aa:Xaa)
    /transl_except= (pos:424..426, aa:Xaa)
    /product= "HPSK protein"
    /note= "Xaa = unknown"
FT FT
PN W09820117-A1.
PD 14-MAY-1998.
PR 31-OCT-1997: U20051.
PR 05-NOV-1996: US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O. GOLL SK.
DR WPI: 98-286933/25.
DR P-PSDB: W60592.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 5: Fig 1A-C: 68pp: English.
CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 871 BP: 166 A; 260 C; 258 G; 184 T;

Query Match 20.08; Score 291.8; DB 1; Length 871;
Best Local Similarity 80.28; Pred. No. 2.7e-55;
Matches 389; Conservative 0; Mismatches 13; Indels 83; Gaps 1;

OY 25 AGTAGTCACAGACTCTACACCATCGGGGCTGCGCTGCACAGTCTTGAGCCGACCAAG 84
    |||||
Db 224 ACTTTTCCAAACTCTCCACACATCGGGGCTGCGCTGCACAGTCTTGAGCCGACCAAG 283

OY 85 AGCAGGAGGAGCAGATGTTGAGGAGGAGCCTCTCGTAGGAGCCACCAAGTAAACAAGAC 144
    |||||
Db 284 AGCAGGAGGAGCAGATGTTGAGGAGGAGCCTCTCGTAGGAGCCACCAAGTAAACAAGAC 343

OY 145 CTTGCTGCTAGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCAGAGTGTGACA 204
    |||||
Db 344 CTTGCTGCTAGACCTCATGCTCATCAAGTTGAGCAATCCGTGTCAGAGTGTGACA 403

OY 205 CCATCCGAGCATCAGCATCTGCTGCGAGTGCCTACCGCGGGGAATCTTGCTCTGTTT 264
    |||||
Db 404 ACATCCCGAGNATCAGCATCTGCTGCGAGTGCCTACCGCGGGGAATCTTGCTCTGTTT 463

OY 265 CTGGCTGGGCTGCTGCTGCGAGCGGTGAGCTCAGGGGTGTGTGTGCTTCAAGA 324
    |||||
Db 464 CTGGCTGGGCTGCTGCTGCGAGC----- 487

OY 325 GGTCTCTGTCCAGTCCGCGGGGGCTGACCCAGAGCTCTGCTCCAGGACAGATGCTTAC 384
    |||||

```

```

Db 487 -----GGCAGATGCTTAC 500

OY 385 CGTGTGAGAGCGGTGAACGCTGGTGTCTGANGAGTCTGCANTAGCTTATGA 444
    |||||
Db 501 CGTGTGAGAGCGGTGAACGCTGGTGTCTGANGAGTCTGCANTAGCTTATGA 560

OY 445 CCCGCTGTACCCACCCANCATGTCTGCGCGCGGAGGAGCCAGACCAAGAGACTCTG 504
    |||||
Db 561 CCCGCTGTACCCACCCAGCATGTCTGCGCGCGGAGGAGCCAGACCAAGAGACTCTG 620

OY 505 CACG 509
    |||||
Db 621 CACG 625

RESULT 11
X41114
ID X41114 standard: cDNA: 402 BP.
AC X41114:
DE 17-JUN-1999 (first entry)
DE Human secreted protein 5', EST SEQ ID NO:58.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; hematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokine; hemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
OS Homo sapiens.
PN W09906548-A2.
PS 11-FEB-1999.
PR 31-JUL-1998: IB1222.
PR 01-AUG-1997: US-905135.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI: 99-153778/13.
DR P-PSDB: Y12281.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,
PT kidney, lung, umbilical cord, placenta and colon tissue
PS Claim 1: Page 205: 824pp: English.
CC X41094 to X41347 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12261 to Y12514,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, hematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, hemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 402 BP: 80 A; 125 C; 121 G; 75 T;

Query Match 19.88; Score 289.4; DB 1; Length 402;
Best Local Similarity 99.08; Pred. No. 7.6e-55;
Matches 290; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 GTACCGGACACAGTCTTCCAGAGTAGTGACAGCTCTACACATCGGGGCTGAGCC 61
    |||||
Db 108 GTACCGGACACAGTCTTCCAGAGTAGTGACAGCTCTACACATCGGGGCTGAGCC 167

OY 62 GCACAGTCTTAGGCGCGACCAAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 121
    |||||
Db 168 GCACAGTCTTAGGCGCGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 227

OY 122 AGGACCCAGATGATCAAGACCTTGTGCTGATGATGATGATGATGATGATGATGATG 181

```

```
Db 228 ACGGACACCCAGAGTACACAGACCTTGTCTGCTAACGACCTCATGCTCAAGTTGGA 287
|||||
QY 182 CGAATCGGTCCGAGCTGACACCATCCGAGCATGAGTCTCGAGAGCCCTAC 241
|||||
Db 288 CGAATCGGTCCGAGCTGACACCATCCGAGCATGAGTCTCGAGAGCCCTAC 347
|||||
QY 242 CGGCGGGAACCTTGTCTGCTGCTGGGTCTGCTGGCGAAGCGTGAGC 294
|||||
Db 348 CGGCGGGAACCTTGTCTGCTGCTGGGTCTGCTGGCGAAGCGTGAGC 400

RESULT 12
V58648 standard; cDNA; 1119 BP.
V58648
ID V58648;
AC 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE14.
KW Prostate tumour specific gene; human; prostate cancer; detection;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 34..528
FT FT /*tag= a
FT PN MO9837418-A2.
FT PD 27-AUG-1998.
FT PF 25-FEB-1998; U03690.
FT PR 09-FEB-1998; US-904809.
FT PR 25-FEB-1997; US-806596.
FT PR 01-AUG-1997; US-904809.
FT PA (CORI-) CORIXA CORP.
FT PI Dillion DC, Xu J;
FT DR WPI: 98-480805/41.
FT DR P-PSDB: W69389.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Claim 1; Page 116-117, 141pp; English.
CC This sequence represents a human prostate tumour specific gene, and can
CC be used in the method of the invention. The method is for detecting
CC prostate cancer comprising contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC encoded by this sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancer. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 17.6%; Score 257.4; DB 1; Length 1119;
Best Local Similarity 97.8%; Pred. No. 9.6e-48;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 AGTGAAGTCAGAGCTCTTAACCATCGGCTGGGCTGACAGTCTTGAGCCGACCAAG 84
|||||
Db 98 ACTGTTCCAGAACTCTTACACCATCGGCTGGGCTGACAGTCTTGAGCCGACCAAG 157
|||||
QY 85 AGCCAGGAGCCAGATGCTGAGAGCCAGCCTCTCCGTACGCAACCCAGATCAACAGAC 144
|||||
Db 158 AGCCAGGAGCCAGATGCTGAGAGCCAGCCTCTCCGTACGCAACCCAGATCAACAGAC 217
|||||
QY 145 CCTTGTCTGCTAAGACCTATGCTCATCAAGTTGAGAGATCCGTCGAGTCTGACA 204
|||||
Db 218 CCTTGTCTGCTAAGACCTATGCTCATCAAGTTGAGAGATCCGTCGAGTCTGACA 277
|||||
QY 205 CCATCCGAGCATCAGATGCTTGCAGATGCCCTACCGGGGGAACCTTTCCTGCTTT 264
|||||
Db 278 CCATCCGAGCATCAGATGCTTGCAGATGCCCTACCGGGGGAACCTTTCCTGCTTT 337
|||||
QY 265 CTGGCTGGGCTGCTGGCGAAGCGTG 291
|||||
Db 338 CTGGCTGGGCTGCTGGCGAAGCGATG 364
```

```
RESULT 13
V61253 standard; cDNA; 1119 BP.
V61253
ID V61253;
AC 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE14.
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 34..528
FT FT /*tag= a
FT PN MO9837418-A2.
FT PD 27-AUG-1998.
FT PF 25-FEB-1998; U03492.
FT PR 09-FEB-1998; US-020956.
FT PR 25-FEB-1997; US-806099.
FT PR 01-AUG-1997; US-904804.
FT PA (CORI-) CORIXA CORP.
FT PI Dillion DC, Xu J;
FT DR WPI: 98-609886/51.
FT DR P-PSDB: W71873.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 108-109; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic portion
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 17.6%; Score 257.4; DB 1; Length 1119;
Best Local Similarity 97.8%; Pred. No. 9.6e-48;
Matches 261; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 25 AGTGAAGTCAGAGCTCTTAACCATCGGCTGGGCTGACAGTCTTGAGCCGACCAAG 84
|||||
Db 98 ACTGTTCCAGAACTCTTACACCATCGGCTGGGCTGACAGTCTTGAGCCGACCAAG 157
|||||
QY 85 AGCCAGGAGCCAGATGCTGAGAGCCAGCCTCTCCGTACGCAACCCAGATCAACAGAC 144
|||||
Db 158 AGCCAGGAGCCAGATGCTGAGAGCCAGCCTCTCCGTACGCAACCCAGATCAACAGAC 217
|||||
QY 145 CCTTGTCTGCTAAGACCTATGCTCATCAAGTTGAGAGATCCGTCGAGTCTGACA 204
|||||
Db 218 CCTTGTCTGCTAAGACCTATGCTCATCAAGTTGAGAGATCCGTCGAGTCTGACA 277
|||||
QY 205 CCATCCGAGCATCAGATGCTTGCAGATGCCCTACCGGGGGAACCTTTCCTGCTTT 264
|||||
Db 278 CCATCCGAGCATCAGATGCTTGCAGATGCCCTACCGGGGGAACCTTTCCTGCTTT 337
|||||
QY 265 CTGGCTGGGCTGCTGGCGAAGCGTG 291
|||||
Db 338 CTGGCTGGGCTGCTGGCGAAGCGATG 364
```

```
RESULT 14
T18325/C standard; DNA; 24026 BP.
T18325
ID T18325;
AC 05-JUN-1996 (first entry)
DE BRCA1, human breast and ovarian cancer predisposing gene.
KW BRCA1; breast cancer; ovarian cancer; predisposing gene;
KW susceptibility gene; diagnosis; prognosis; gene therapy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Intron 1..55
FT Intron /*tag= a
FT Exon 56..155
FT Exon /*tag= b
FT Intron 156..1512
FT Exon /*tag= c
FT Exon 1513..1611
```



FT	Intron	/tag- d
FT		1612. .2206
FT		/tag- e
FT		/note= "n at 1925-1937 represent an indefinite
FT		interval within the Intron"
FT	exon	2207. .2260
FT		/tag- f
FT		2261. .2677
FT		/tag- g
FT		/note= "n at 2569-2581 represent an indefinite
FT		interval within the Intron"
FT	exon	2678. .2788
FT		/tag- h
FT		2789. .3328
FT		/tag- i
FT		/note= "n at 3063-3075 represent an indefinite
FT		interval within the Intron"
FT	exon	3329. .3406
FT		/tag- j
FT		3407. .3813
FT		/tag- k
FT		/note= "n at 3598-3610 represent an indefinite
FT		interval within the Intron"
FT	exon	3814. .3902
FT		/tag- l
FT		3903. .4224
FT		/tag- m
FT		/note= "n at 4076-4088 represent an indefinite
FT		interval within the Intron"
FT	Intron	4225. .4364
FT		/tag- n
FT	exon	4365. .6571
FT		/tag- o
FT		/note= "n at 4602-4614 represent an indefinite
FT		interval within the Intron"
FT	exon	6572. .6677
FT		/tag- p
FT		6678. .9163
FT		/tag- q
FT	exon	9164. .9207
FT		/tag- r
FT	Intron	9208. .10530
FT		/tag- s
FT		10531. .10607
FT	exon	/tag- t
FT		10608. .11597
FT		/tag- u
FT		/note= "n at 11383-11396 represent an indefinite
FT		interval within the Intron"
FT	exon	11598. .15023
FT		/tag- v
FT	Intron	15024. .15424
FT		/tag- w
FT	exon	15425. .15511
FT		/tag- x
FT	Intron	15512. .15952
FT		/tag- y
FT		/note= "n at 15647-15659 represent an indefinite
FT		interval within the Intron"
FT	exon	15953. .16126
FT		/tag- z
FT	Intron	16127. .16565
FT		/tag- aa
FT		/note= "n at 16370-16382 represent an indefinite
FT		interval within the Intron"
FT	exon	16566. .16692
FT		/tag- ab
FT	Intron	16693. .17535
FT		/tag- ac
FT		/note= "n at 17290-17302 represent an indefinite
FT		interval within the Intron"
FT	exon	17536. .17726
FT		/tag- ad

---

FT	Intron	17727. .18416
FT		/tag- ae
FT		/note= "n at 18299-18312 represent an indefinite
FT		interval within the Intron"
FT	exon	18417. .18787
FT		/tag- af
FT	Intron	18788. .19298
FT		/tag- ag
FT		/note= "n at 18952-18964 represent an indefinite
FT		interval within the Intron"
FT	exon	19299. .19386
FT		/tag- ah
FT	Intron	19387. .20190
FT		/tag- ai
FT		/note= "n at 19887-19899 represent an indefinite
FT		interval within the Intron"
FT	exon	20191. .20267
FT		/tag- aj
FT	Intron	20268. .21094
FT		/tag- ak
FT		/note= "n at 20767-20779 represent an indefinite
FT		interval within the Intron"
FT	exon	21095. .21135
FT		/tag- al
FT	Intron	21136. .21583
FT		/tag- am
FT		/note= "n at 21341-21353 represent an indefinite
FT		interval within the Intron"
FT	exon	21584. .21667
FT		/tag- an
FT	Intron	21668. .22233
FT		/tag- ao
FT		/note= "n at 21921-21933 represent an indefinite
FT		interval within the Intron"
FT	exon	22234. .22288
FT		/tag- ap
FT	Intron	22289. .22832
FT		/tag- aq
FT		/note= "n at 22567-22579 represent an indefinite
FT		interval within the Intron"
FT	exon	22833. .22906
FT		/tag- ar
FT	Intron	22907. .23287
FT		/tag- as
FT		/note= "n at 23050-23062 represent an indefinite
FT		interval within the Intron"
FT	exon	23288. .23348
FT		/tag- at
FT	Intron	23349. .23698
FT		/tag- au
FT		/note= "n at 23580-23592 represent an indefinite
FT		interval within the Intron"
FT	exon	23699. .24026
FT		/tag- av
FT		2725
FT		/tag- aw
FT		/note= "polymorphic site"
FT		3653
FT		/tag- ax
FT		/note= "polymorphic site"
FT		4391
FT		/tag- ay
FT		/note= "polymorphic site"
FT		4392
FT		/tag- az
FT		/note= "polymorphic site"
FT		6823
FT		/tag- ba
FT		/note= "polymorphic site"
FT		9106
FT		/tag- bb
FT		/note= "polymorphic site"
FT		9207

misc\_feature

```

FT      /tag- bc
FT      /note- "polymorphic site"
FT      9376
FT      /tag- bd
FT      /note- "polymorphic site"
FT      11908
FT      /tag- be
FT      /note- "polymorphic site"
FT      11994
FT      /tag- bf
FT      /note- "polymorphic site"
FT      12952
FT      /tag- bg
FT      /note- "polymorphic site"
FT      13004
FT      /tag- bh
FT      /note- "polymorphic site"
FT      13009
FT      /tag- bi
FT      /note- "polymorphic site"
FT      13048
FT      /tag- bj
FT      /note- "polymorphic site"
FT      13238
FT      /tag- bk
FT      /note- "polymorphic site"
FT      13448
FT      /tag- bl
FT      /note- "polymorphic site"
FT      13539
FT      /tag- bm
FT      /note- "polymorphic site"
FT      13951
FT      /tag- bn
FT      /note- "polymorphic site"
FT      14041
FT      /tag- bo
FT      /note- "polymorphic site"
FT      14046

```

Query Match 15.5%; Score 226.2; DB 1; Length 24026;

Best Local Similarity 87.0%; Pred. No. 1.3e-40;

Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

```

Oy      1162 GCTGGGACAGGCGCTCATGCTGAATCCAGCACTTGGAGGC-GAGGACGACAGT 1220
DB      10357 GCTGGGACAGGCGCTCATGCTGAATCCAGCACTTGGAGGC-GAGGCGGTGGAC 10298
Oy      1221 CACTTGAGGTAGAGATTCAAGACCAGCTGGCCAAATGTGAATCCTGTCTACTA 1280
DB      10297 CACTTGAGGTATGATGAATCAAGACCAGCTGGCCAAATGTGAATCCTGTCTACTA 10238
Oy      1281 AAAATATCAAAAGTTAGCTGGATATGCTGGCAGCGCCTTAATCCAGCTACTTGGAGG 1340
DB      10237 AAAATATCAAAAGTTAGCTGGATATGCTGGCAGCGCCTTAATCCAGCTACTTGGAGG 10178
Oy      1341 CTGAGGACGAGAGATTGCTGTGAATATGGAGGACAGGTTGAAGTGTGATGATCAGAC 1400
DB      10177 CTGAGGACGAGAGATTGCTGTGAATATGGAGGACAGGTTGCGAGAGGTGATGATCAGAC 10118
Oy      1401 CACTTACTCCAGCTGGGACACAGAGTAAAGACTGTCTCAAAAAA 1459
DB      10117 CTCTGTACTCCAGCTGGGACACAGAGTCTGTCTCAAAAAA 10059

```

RESULT 15  
117455/c  
ID 117455 standard; cDNA; 24025 BP.  
AC 117455;  
DT 07-OCT-1996 (first entry)  
DE Mutated BRCA1 genomic sequence from sample set MSKCC family 19921.  
KW Cancer therapy; breast and ovarian cancer predisposing gene; immunogen;  
mitbody-production; germline alteration; probe; lesion neoplasia; human;

```

KW      gene therapy: protein replacement therapy; protein mimetic; BRCA1; ds.
OS      Homo sapiens.
PH      Key
FT      exon
FT      256..355
FT      /tag- a
FT      /note- "exon 1"
FT      356..1512
FT      /tag- b
FT      /note- "intron 1"
FT      1295
FT      /tag- c
FT      /note- "known polymorphic site"
FT      1513..1611
FT      /tag- d
FT      /note- "exon 2"
FT      1612..2206
FT      /tag- e
FT      /note- "intron 2"
FT      1925..1937
FT      /tag- f
FT      /note- "indefinite interval within intron 2"
FT      2141
FT      /tag- g
FT      /note- "known polymorphic site"
FT      2207..2260
FT      /tag- h
FT      /note- "exon 3"
FT      2261..2677
FT      /tag- i
FT      /note- "intron 3"
FT      2569..2581
FT      /tag- j
FT      /note- "indefinite interval within intron 3"
FT      2678..2788
FT      /tag- k
FT      /note- "exon 4"
FT      2725
FT      /tag- l
FT      /note- "known polymorphic site"
FT      2789..3328
FT      /tag- m
FT      /note- "intron 4"
FT      3063..3075
FT      /tag- n
FT      /note- "indefinite interval within intron 4"
FT      3329..3406
FT      /tag- o
FT      /note- "exon 5"
FT      3407..3813
FT      /tag- p
FT      /note- "intron 5"
FT      3598..3610
FT      /tag- q
FT      /note- "indefinite interval within intron 5"
FT      3653
FT      /tag- r
FT      /note- "known polymorphic site"
FT      3814..3902
FT      /tag- s
FT      /note- "exon 6"
FT      3903..4224
FT      /tag- t
FT      /note- "intron 6"
FT      4223
FT      /tag- u
FT      /note- "site of 1 nucleotide deletion"
FT      4076..4088
FT      /tag- v
FT      /note- "indefinite interval within intron 6"
FT      4225..4364
FT      /tag- w
FT      /note- "exon 7"
FT      4365..6571
FT      intron

```

```

FT      /tag- x
FT      /note- "intron 7"
FT      4391.4392
FT      /tag- y
FT      /note- "known polymorphic site"
FT      4602.4614
FT      /tag- z
FT      /note- "indefinite interval within intron 7"
FT      6538
FT      /tag- aa
FT      /note- "known polymorphic site"
FT      6572.6677
FT      /tag- ab
FT      /note- "exon 8"
FT      6678.9163
FT      /tag- ac
FT      /note- "intron 8"
FT      6823
FT      /tag- ad
FT      /note- "known polymorphic site"
FT      9106
FT      /tag- ae
FT      /note- "known polymorphic site"
FT      9184.9209
FT      /tag- af
FT      /note- "exon 9"
FT      9207
FT      /tag- ag
FT      /note- "known polymorphic site"
FT      9210.10530
FT      /tag- ah
FT      /note- "intron 9"
FT      9376
FT      /tag- ai
FT      /note- "known polymorphic site"
FT      10531.10607
FT      /tag- aj
FT      /note- "exon 10"
FT      10608.11597
FT      /tag- ak
FT      /note- "intron 10"
FT      11384.11396
FT      /tag- al
FT      /note- "indefinite interval within intron 10"
FT      11598.15023
FT      /tag- am
FT      /note- "exon 11"
FT      11908
FT      /tag- an
FT      /note- "known polymorphic site"
FT      11994
FT      /tag- ao
FT      /note- "known polymorphic site"
FT      12952
FT      /tag- ap
FT      /note- "known polymorphic site"
FT      13004
FT      /tag- aq
FT      /note- "known polymorphic site"
FT      13009
FT      /tag- ar
FT      /note- "known polymorphic site"
FT      13048
FT      /tag- as
FT      /note- "known polymorphic site"
FT      13338
FT      /tag- at
FT      /note- "known polymorphic site"
FT      13448
FT      /tag- au
FT      /note- "known polymorphic site"
FT      13539
FT      /tag- av

```

```

FT      /note- "known polymorphic site"
FT      13951
FT      /tag- aw
FT      /note- "known polymorphic site"
FT      14041
FT      /tag- ax
FT      /note- "known polymorphic site"
FT      14046
FT      /tag- ay
FT      /note- "known polymorphic site"
FT      14475
FT      /tag- az
FT      /note- "known polymorphic site"
FT      14874
FT      /tag- ba
FT      /note- "known polymorphic site"
FT      14891
FT      /tag- bb
FT      /note- "known polymorphic site"
FT      14966
FT      /tag- bc
FT      /note- "known polymorphic site"
FT      15024.15424
FT      /tag- bd
FT      /note- "intron 11"
FT      15284
FT      /tag- be
FT      /note- "known polymorphic site"
FT      15425.15511
FT      /tag- bf
FT      /note- "exon 12"
FT      15512.15952
FT      /tag- bg
FT      /note- "intron 12"
FT      15647.15659
FT      /tag- bh
FT      /note- "indefinite interval within intron 12"
FT      15953.16126
FT      /tag- bi
FT      /note- "exon 13"
FT      16077
FT      /tag- bj
FT      /note- "known polymorphic site"
FT      16127.16565
FT      /tag- bk
FT      /note- "intron 13"
FT      16243
FT      /tag- bl

```

```

Query Match      15.5%; Score 226.2; DB 1; Length 24025;
Best Local Similarity 87.0%; Pred. No. 1.3e+40;
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY      1162 GCTGGCAGGCGTGCATCCCTGTAATCCACAGCTTTGGGAGGC-GAGGCAGCAGAT 1220
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB      10356 GCTGGCAGGCGTGCATCCCTGTAATCCACAGCTTTGGGAGGCAGCAGAT 10297

QY      1221 CACTTGAGGTAAGGAGTTCAAGACCAAGCTGGCCAAATGCTGCTGTACTA 1280
          ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB      10296 CACTTGAGGTCATGATCAAGACCAAGCTGGCCAAATGCTGCTGTACTA 10237

QY      1281 AAAATCAAAAGTACTGATATGTTGGCAGGCGCTGTAATCCAGCTACTGGAGG 1340
          ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB      10236 AAAATCAAAAGTACTGATATGTTGGCAGGCGCTGTAATCCAGCTACTGGAGG 10177

QY      1341 CTGAGCAGAGAAATTCCTGAATATGGAGCAGAGTTGAAGTTGAGATCACAC 1400
          ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB      10176 CTGAGCAGAGAAATTCCTGAATATGGAGCAGAGTTGAGATCACAC 10117

QY      1401 CACTTACTCCAGCTGGGCGACAGAGTAAGACTGTCTCAAAAAAAAAAAAAA 1459
          ||||||| ||||||| ||||||| ||||||| ||||||| ||
DB      10116 CTTCTGTAATCCAGCTGGGCGACAGAGTAAGACTGTCTCAAAAAAAAAAAAAA 10058

```

Wed Sep 29 14:27:21 1999

us-09-030-606-174.rng

Page 12

Search completed: September 28, 1999, 12:28:26  
Job time: 5326 sec

---

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:33:30 ; Search time 161.06 Seconds  
(without alignments)  
830.474 Million cell updates/sec

Title: US-09-030-606-174

Sequence: 1459  
1 GCTCAGCCGACACTGTTC.....TCAAAAAAAAAAAAAAAAAA 1459

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database: Issued Patents, NA.\*

1: /cgn2\_6/ptodata/2/1na/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/1na/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/1na/5C.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/1na/5D.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/1na/PTUS9.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/1na/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	291.8	20.0	871	3	US-08-744-026-2	Sequence 2, Appl1
2	226.2	15.5	6769	2	US-08-480-784-20	Sequence 20, Appl1
3	226.2	15.5	6769	2	US-08-483-553-20	Sequence 20, Appl1
4	226.2	15.5	6769	2	US-08-487-002-20	Sequence 20, Appl1
5	226.2	15.5	6769	3	US-08-483-554B-20	Sequence 20, Appl1
6	226.2	15.5	6769	3	US-08-488-011B-20	Sequence 20, Appl1
7	226.2	15.5	6769	5	PCT-US95-10202-20	Sequence 20, Appl1
8	226.2	15.5	6769	5	PCT-US95-10203-20	Sequence 20, Appl1
9	226.2	15.5	6769	5	PCT-US95-10220-20	Sequence 20, Appl1
10	219.2	15.0	7676	3	US-08-451-777A-7	Sequence 7, Appl1
11	219.2	15.0	7676	4	US-08-998-208-7	Sequence 7, Appl1
12	219.2	15.0	7676	4	US-08-998-208-7	Sequence 7, Appl1
13	219.2	15.0	7676	4	US-08-998-208-7	Sequence 7, Appl1
14	213.4	14.6	26764	1	US-08-370-975B-1	Sequence 1, Appl1
15	213.4	14.6	20303	1	US-08-370-975B-6	Sequence 6, Appl1
16	212.2	14.5	4823	5	PCT-US94-08806-28	Sequence 28, Appl1
17	212.2	14.5	4823	5	PCT-US95-01829-5	Sequence 5, Appl1
18	212.2	14.5	4823	5	PCT-US95-16626-5	Sequence 5, Appl1
19	211.8	14.5	22481	5	PCT-US95-07201-43	Sequence 43, Appl1
20	211.2	14.5	246240	4	US-08-724-394A-20	Sequence 20, Appl1
21	211.2	14.5	246240	4	US-08-724-394A-21	Sequence 21, Appl1
22	211.2	14.5	246240	4	US-08-724-394A-22	Sequence 22, Appl1
23	210.6	14.4	17327	1	US-07-906-871-15	Sequence 15, Appl1
24	209.8	14.4	2461	3	US-08-832-883-3	Sequence 3, Appl1
25	209.8	14.4	2461	3	US-08-832-883-3	Sequence 3, Appl1
26	209.8	14.4	1811	3	US-08-848-252-1	Sequence 113, Appl1
27	207.8	14.2	1988	4	US-08-257-963B-11	Sequence 11, Appl1
28	207.8	14.2	1988	4	US-08-257-963B-11	Sequence 11, Appl1
29	207.6	14.2	7210	4	US-08-257-963B-10	Sequence 10, Appl1
30	207.6	14.2	7210	5	PCT-US95-07201-10	Sequence 10, Appl1
31	207.6	14.2	22481	5	PCT-US95-07201-43	Sequence 43, Appl1
32	206.8	14.2	282	1	US-08-133-629-8	Sequence 8, Appl1
33	203.6	14.0	2472	2	US-08-428-742-1	Sequence 1, Appl1
34	203.2	13.9	7620	1	US-07-767-135-1	Sequence 1, Appl1
35	203.2	13.9	7620	1	US-07-841-652-1	Sequence 1, Appl1
36	202.4	13.9	246240	4	US-08-724-394A-20	Sequence 20, Appl1
37	202.4	13.9	246240	4	US-08-724-394A-21	Sequence 21, Appl1

## ALIGNMENTS

38	202.4	13.9	246240	4	US-08-724-394A-22	Sequence 22, Appl1
39	201.4	13.8	3176	4	US-08-910-733-17	Sequence 17, Appl1
40	199.4	13.7	8392	1	US-08-080-255-6	Sequence 6, Appl1
41	199.4	13.7	8392	5	PCT-US93-05857-6	Sequence 5, Appl1
42	198.4	13.6	8174	1	US-07-914-281-5	Sequence 5, Appl1
43	198.4	13.6	8174	1	US-08-393-246-5	Sequence 5, Appl1
44	198.4	13.6	11531	1	US-08-068-945A-1	Sequence 1, Appl1
45	198.4	13.6	3373	1	US-08-273-411-2	Sequence 2, Appl1

RESULT 1  
US-08-744-026-2  
Sequence 2, Application US/08744026

Patent No. 5786148  
GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
APPLICANT: Goll, Surya K.

TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
TITLE OF INVENTION: KALIKREIN

NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:

ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive

CITY: Palo Alto  
STATE: CA

COUNTRY: US  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: diskette

OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,026

FILING DATE: Herewith  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:

FILING DATE:  
ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0154 US  
TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:

LENGTH: 871 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear

IMMEDIATE SOURCE:  
LIBRARY:

CLONE: Consensus  
US-08-744-026-2

Query Match 20.0% Score 291.8; DB 3; Length 871;  
Best Local Similarity 80.2% Pred. No. 1.5e-64;  
Matches 389; Conservative 0; Mismatches 13; Indels 83; Gaps 1;

QY	25	AGTAGTGACAGGCTCTACACATCGGGCTGGGCTGCACAGTCTGAGGCGCAGCAAG	84
DB	224	ACTTTTCAGAAATCTCTACACCATCGGGCTGGGCTGCACAGTCTGAGGCGCAGCAAG	283
QY	85	AGCAGGAGGAGCAGATGATGAGGAGCAGCTCTCGTAGGCGCAGCAGATACAGAC	144
DB	284	AGCAGGAGGAGCAGATGATGAGGAGCAGCTCTCGTAGGCGCAGCAGATACAGAC	343

QY	145	CTTGTCGTGCTTAACGACCTCATGCTCAACAGTTGGAGCAATCCGTGCGAGTGTGACA	204
Db	344	CTTGTCGTGCTTAACGACCTCATGTCATCATCACTTGGAGCAATCCGTGTCCGAGTCTACA	403
QY	205	CCATCCGGAGCATCAGCATTTGCTCTCGCAGTCCCTACCGCGGGGAACTCTTGCTCTGTTT	264
Db	404	ACATCCGGAGNATCAGCATTTGTTGGAGTCCCTACCGGGGGAACTTTTGCTCTGTTT	463
QY	265	CTGCGTGGGGGTCTGCTGGCGCAACGCTGAGCTCACGGGTGTGTCTGCCCTCTTCAAAGA	324
Db	464	CTGCGTGGGGGTCTGCTGGCGCAAC-----	487
QY	325	GGTCTCTGGCCAGTCGGGGGGGCTGACCCAGACCTCTGCGTCCAGCGCAAGATCCCTAC	384
Db	487	-----	500
QY	385	CGTGTGTCAGTGCCTGAACTGTGCGTGGTGTCTGANGAGTCTGCANTAACTCTATGA	444
Db	501	CGTGTGTCAGTGCCTGAACTGTGCGTGGTGTCTGAGAGAGTCTGCAATGAAGCTCTATGA	560
QY	445	CCCCGTGTACCAACCCCAACATGTTCTGTGCGCGGGCGGAGGCAAGACCAAGAGAAGTCTCTG	504
Db	561	CCCCGTGTACCAACCCCAACATGTTCTGTGCGCGGGCGGAGGCAAGACCAAGAGAAGTCTCTG	620
QY	505	CAACG 509	
Db	621	CAACG 625	

RESULT 2  
 US-08-480-784-20/c  
 Sequence 20, Application US/08480784  
 Patent No. 5693473  
 GENERAL INFORMATION:  
 APPLICANT: Skolnick, Mark H.  
 APPLICANT: Goldgar, David E.  
 APPLICANT: Miki, Yoshio  
 APPLICANT: Swenson, Jeff  
 APPLICANT: Kamb, Alexander  
 APPLICANT: Harshman, Keith D.  
 APPLICANT: Shattuck-Eidens, Donna M.  
 APPLICANT: Tavtigian, Sean V.  
 APPLICANT: Wiseman, Roger W.  
 TITLE OF INVENTION: Futreal, P. Andrew  
 TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Venable, Baetjer, Howard & Civiletti, LLP  
 STREET: 1201 New York Avenue, N.W., Suite 1000  
 CITY: Washington  
 STATE: DC  
 COUNTRY: USA  
 ZIP: 20005  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/480,784  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/409,305  
 FILING DATE: 24-MAR-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/348,824  
 FILING DATE: 29-NOV-1994  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/308,104  
 FILING DATE: 16-SEP-1994  
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 28,957  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHEITICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens

Query Match	15.5%;	Score 226.2;	DB 2;	Length 6769;
Best Local Similarity	87.0%;	Pred. No. 9.5e-48;		
Matches 260; Conservative	0;	Mismatches 38;	Indels 1;	Gaps 1

QY	1162	GCTGGGACGGGTGGCATCTCCGTAAATCCAGACATTTGGGAGGC - GAGCGAGGCAGT	1220
Db	5743	GCATGGACCGGTGGCTCAACCCCTCAATTCACACACTTTGGGAGGCTAGCGGGGTGGAC	5688
QY	1221	CACTTGAGTTAAGAGATTCAAGACCAAGCCTGGCCAAAATGTGTAAATCTGTCTACTA	1280
Db	5683	CACTTGAGGTTCATGAGTTCAAGACCAAGCCTGGCCAACTGTGTAAATCCCATCTCTACTA	5622
QY	1281	AAATTAACAAAAGTAGTGGTGAATGTGGACGAGCGCGCTTAATCCAGCTACTTTGGGAGG	1340
Db	5623	AAATTAACAAAATTTAGCCGGGGTGGTGGCGCGCCCTTAAGTCCAGCTACTTAAGAGAGG	5566
QY	1341	CTGAGCGAGAGAAATGCTTGAATATGGAGGCGAGAGGTTGAAGTGTGAGATCACAC	1400
Db	5563	CTGAGCGAGAGATATGGCTTCAACCTGGGAGGCGAGAGCTTGGGAGAGGTGAGATCACAC	5504
QY	1401	CACATATCTCTACACTGGGCGCAACAGTGTAAGCTGTCTCAAAAAAAAAAAAAAAAAAAAA	1459
Db	5503	CTCTGTACTCCAGCTGGGGCAACGAGCAAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAA	5445

RESULT 3  
 US-08-483-553--20/c  
 Sequence 20 Application US/08483553  
 Patent No. 5/09399  
 GENERAL INFORMATION:  
 APPLICANT: Skolnick, Mark H.  
 APPLICANT: Goldfar, David E.  
 APPLICANT: Miki, Yoshio  
 APPLICANT: Swenson, Jeff  
 APPLICANT: Kamd, Alexander  
 APPLICANT: Harshman, Keith D.  
 APPLICANT: Shattuck-Eidens, Donna M.  
 APPLICANT: Tavtigian, Sean V.  
 APPLICANT: Wiseman, Roger W.  
 APPLICANT: Futreal, P. Andrew  
 TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
 TITLE OF INVENTION: Susceptibility Gene  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
 STREET: 1201 New York Avenue, N.W., Suite 1000  
 CITY: Washington

STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,553  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,824  
FILING DATE: 29-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/308,104  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-483-553-20

Query Match 15.5% Score 226.2; DB 2; Length 6769;  
Best Local Similarity 87.0%; Pred. No. 9.5e-48;  
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;  
DB 1162 GCTGGCAGGCTGCTCCTGCTTAATCCAGCAGCTTTGGAGGC-GAGCAGGCAGAT 1220  
|||||  
DB 5743 GCTGGCAGGCTGCTCCTGCTTAATCCAGCAGCTTTGGAGGCAGGCGGTGGAC 5684  
|||||  
DB 1221 CACTTAGGTAGAGGTTCAAGACAGCCTGCGCAAAATGTGAATCTCTCTGTA 1280  
|||||  
DB 5683 CACTTAGGTAGATAGTCAAGACAGCCTGCGCAACATGTGAACCCCTCTCTACTA 5624  
|||||  
DB 1281 AAAATACAAAAGTATAGTATGTTGGCAGGCGCTGTAATCCAGCTACTGGGAGG 1340  
|||||  
DB 5623 AAAATACAAAATTAATGCGGCTGTTGGCGCGGCTGTATGCTCCAGCTATGAGAGG 5564  
|||||  
DB 1341 CTGAGCAGAGAAATGCTTGAATATGGAGCAGAGGTTGAAGTGAATGATCAAC 1400  
|||||  
DB 5563 CTGAGCAGAGGATATGCTTCAACCTGGAGCAGAGCTGCGGAGGAGTGAATCAAC 5504  
|||||  
DB 1401 CACTTACTCAGCTGGGCGCAACAGATGATCTGTCTCAAAAAAAAAAAAAA 1459  
|||||  
DB 5503 CTCTGACTCCAGCTGGCGCAACAGAGCAAGACTCTGTCTCAAAAAAAAAAAAAA 5445  
|||||

RESULT 4  
US-08-487-002-20/c  
Sequence 20. Application US/08487002  
Patent No. 5710001  
GENERAL INFORMATION:  
APPLICANT: Shattuck-Eidens, Donna M.  
APPLICANT: Simard, Jacques  
APPLICANT: Eml, Mitsuru  
APPLICANT: Nakamura, Yusuke  
APPLICANT: Dutocher, Francine  
TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,002  
FILING DATE:  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,824  
FILING DATE: 29-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/308,104  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-487-002-20

Query Match 15.5% Score 226.2; DB 2; Length 6769;  
Best Local Similarity 87.0%; Pred. No. 9.5e-48;  
Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;  
DB 1162 GCTGGCAGGCTGCTCCTGCTTAATCCAGCAGCTTTGGAGGC-GAGCAGGCAGAT 1220  
|||||  
DB 5743 GCTGGCAGGCTGCTCCTGCTTAATCCAGCAGCTTTGGAGGCAGGCGGTGGAC 5684  
|||||





PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,824  
FILING DATE: 29-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/308,104  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109347-09  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-08-488-011B-20

Query Match 15.5% Score 226.2; DB 3; Length 6769;  
Best Local Similarity 87.0%; Pred. No. 9.5e-48;

Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

1162 GCTGGGAGAGGCTGCTATGCTGTAATCCAGACACTTTGGGAGGCG-GAGGACAGCAGAT 1220

5743 GCTGGGAGAGGCTGCTATGCTGTAATCCAGACACTTTGGGAGGCGGAGGAGGAC 5684

1221 CACTTGAGGTAAAGATTCAAGACAGCAGCTGCGCAAAATGTTGAATCTCTCTACTA 1280

5683 CACTTGAGGTAAAGATTCAAGACAGCAGCTGCGCAAAATGTTGAATCTCTCTACTA 5624

1281 AAATACAAAGATTAGTATGATGATGAGGAGGCGCTGTAATCCAGACTCTGGGAGG 1340

5623 AAATACAAAGATTAGTATGATGATGAGGAGGCGCTGTAATCCAGACTCTGGGAGG 5564

1341 CTGAGGAGAGGAATGCTGTAATATGAGGAGCAGAGGTTGAAGTGAATGATGATCAGC 1400

5563 CTGAGGAGAGGAATGCTGTAATATGAGGAGCAGAGGTTGAAGTGAATGATGATCAGC 5504

1401 CACTTACTCAGCTGGGAGCAGACAGATGATGATGATGATGATGATGATGATGATGAT 1459

5503 CTCTGCTCTCAGCTGGGAGCAGACAGATGATGATGATGATGATGATGATGATGATGAT 5445

# RESULT 7

PCT-US95-10202-20/c

Sequence 20, Application PC/TUS9510202

GENERAL INFORMATION:

APPLICANT: Shattuck-Bidens, Donna M.

APPLICANT: Shattuck-Bidens, Donna M.

APPLICANT: Shattuck-Bidens, Donna M.

APPLICANT: Shattuck-Bidens, Donna M.

APPLICANT: Shattuck-Bidens, Donna M.

NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10202  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,824  
FILING DATE: 29-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08-308,104  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
PCT-US95-10202-20

Query Match 15.5% Score 226.2; DB 5; Length 6769;  
Best Local Similarity 87.0%; Pred. No. 9.5e-48;

Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

1162 GCTGGGAGAGGCTGCTATGCTGTAATCCAGACACTTTGGGAGGCG-GAGGACAGCAGAT 1220

5743 GCTGGGAGAGGCTGCTATGCTGTAATCCAGACACTTTGGGAGGCGGAGGAGGAC 5684

1221 CACTTGAGGTAAAGATTCAAGACAGCAGCTGCGCAAAATGTTGAATCTCTCTACTA 1280

5683 CACTTGAGGTAAAGATTCAAGACAGCAGCTGCGCAAAATGTTGAATCTCTCTACTA 5624

1281 AAATACAAAGATTAGTATGATGATGAGGAGGCGCTGTAATCCAGACTCTGGGAGG 1340

5623 AAATACAAAGATTAGTATGATGATGAGGAGGCGCTGTAATCCAGACTCTGGGAGG 5564

1341 CTGAGGAGAGGAATGCTGTAATATGAGGAGCAGAGGTTGAAGTGAATGATGATCAGC 1400

Db	5563	CTGAGCGAGAGTATGGCTTCAACCTGGGAGCGACGTTGGCGAGAGGTGAGATCACAC	5504
Qy	1401	CACATATATCCAGCTGGGGCACAAGAGTAAGACTCTGTCTCAAAAAAAAAAAAAAAAAA	1459
Db	5503	CTCTGTACTCCAGCCTTGGGCCAACAGACAGCAAGCTCTGTCTCAAAAAAAAAAAAAAAAAA	5445

## RESULT 8

```

1 PCT-US95-10203-20/c
2 Sequence 20, Application PC/RUS9510203
3 GENERAL INFORMATION:
4 APPLICANT: Skolnick, Mark H.
5 APPLICANT: Goldgar, David E.
6 APPLICANT: Miki, Yoshio
7 APPLICANT: Swenson, Jeff
8 APPLICANT: Kamb, Alexander
9 APPLICANT: Harshman, Keith D.
10 APPLICANT: Shattuck-Eidens, Donna M.
11 APPLICANT: Tavtigian, Sean V.
12 APPLICANT: Wiseman, Roger W.
13 APPLICANT: Futreal, P. Andrew
14 TITLE OF INVENTION: 17q-Linked Breast and Ovarian Cancer
15 TITLE OF INVENTION: Susceptibility Gene
16 NUMBER OF SEQUENCES: 85
17 CORRESPONDENCE ADDRESSES:
18 ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
19 STREET: 101 New York Avenue, N.W., Suite 1000
20 CITY: Washington
21 STATE: DC
22 COUNTRY: USA
23 ZIP: 20005
24 COMPUTER READABLE FORM:
25 MEDIUM TYPE: Floppy disk
26 COMPUTER: IBM PC compatible
27 OPERATING SYSTEM: PC-DOS/MS-DOS
28 SOFTWARE: PatentIn Release #1.0, Version #1.30
29 CURRENT APPLICATION DATA:
30 APPLICATION NUMBER: PCT/US95/10203
31 FILING DATE:
32 CLASSIFICATION:
33 PRIOR APPLICATION DATA:
34 APPLICATION NUMBER: US
35 FILING DATE: 07-JUN-1995
36 PRIOR APPLICATION DATA:
37 APPLICATION NUMBER: US 08/409,305
38 FILING DATE: 24-MAR-1995
39 PRIOR APPLICATION DATA:
40 APPLICATION NUMBER: US 08/348,824
41 FILING DATE: 29-NOV-1994
42 PRIOR APPLICATION DATA:
43 APPLICATION NUMBER: US 08-308,104
44 FILING DATE: 16-SEP-1994
45 PRIOR APPLICATION DATA:
46 APPLICATION NUMBER: US 08/300,266
47 FILING DATE: 02-SEP-1994
48 PRIOR APPLICATION DATA:
49 APPLICATION NUMBER: US 08/289,221
50 FILING DATE: 12-AUG-1994
51 ATTORNEY/AGENT INFORMATION:
52 NAME: Ihnen, Jeffrey L.
53 REGISTRATION NUMBER: 28,957
54 REFERENCE/DOCKET NUMBER: 24884-109347
55 TELECOMMUNICATION INFORMATION:
56 TELEPHONE: 202-962-4810
57 TELEFAX: 202-962-8300
58 INFORMATION FOR SEQ ID NO: 20:
59 SEQUENCE CHARACTERISTICS:
60 LENGTH: 6769 base pairs
61 TYPE: nucleic acid
62 STRANDEDNESS: double
63 TOPOLOGY: linear
64 MOLECULE TYPE: DNA (genomic)

```

;  
; HYPOTHETICAL: NO  
;  
; ANTI-SENSE: NO  
;  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
;  
PCT-US95-10203-20

Query Match	15.58;	Score 226.2;	DB 5;	Length 6769;
Best Local Similarity	87.08;	Pred. NO. 9.5e-48;		
Matches 260;	Conservative 0;	Mismatches 38;	Indels 1;	Gaps 1;

QY	1162	GCAGGAGAGGGTGGATCATCCCTGTAATCCAGACATTGGGAGGC -GAGCGAGGCAAT	1220
Db	5743	GCTGGGACGCTGGCTCAACCTCTGTAATTCACACATTTTGGGAGGCTGGGGGGTGGAC	5684
QY	1221	CAC TTGAGGTAAAGAGTTTCAAGACCAGCCTGGGCCAAATGATGTAATCCTGTCTACTA	1280
Db	5683	CAC TTGAGGTATAGTTCACAGACCAGCCTGGGCCAATGATGTAATCCATCTCTACTA	5624
QY	1281	AAAAATAAAAGTTACTGTGATATGGTGGCAGGCGCTTATATCCCACTACTTGGGAGG	1340
Db	5623	AAAAATAAAAAATTAAGCGGGGTGTGGTGGCGCGCTGTATCCCACTACTAAGAGG	5564
QY	1341	CTGAGCGAGAGATTCCTGTAATATAGGAGCGACAGATGTGAAGTGGATTGATACAC	1400
Db	5563	CTGAGCGAGAGATATGGCTTCACACCTGGGAGGAGACGTTCCGGAGAGGTGATGATACAC	5504
QY	1401	CACATATCTCCAGCTGGGGCACACAGATGAAGCTCTGTCTCAAAAAAAAAAAAAAAAAA	1459
Db	5503	CTCTGTACTCCAGCTGGGGCACACAGAGACTCTGTCTCAAAAAAAAAAAAAAAAAA	5445

## RESULT

PCT-US95-10220-20/c  
Sequence 20, Application PC/TUS9510220  
GENERAL INFORMATION:  
APPLICANT: Skolnick, Mark H.  
APPLICANT: Goldgar, David E.  
APPLICANT: Miki, Yoshio  
APPLICANT: Swenson, Jeff  
APPLICANT: Kamb, Alexander  
APPLICANT: Harshman, Keith D.  
APPLICANT: Shattuck-Eldens, Donna M.  
APPLICANT: Tavtigian, Sean V.  
APPLICANT: Wiseman, Roger W.  
APPLICANT: Futreal, P. Andrew  
TITLE OF INVENTION: Method for Diagnosing a  
TITLE OF INVENTION: Predisposition for Breast and Ovarian Cancer  
NUMBER OF SEQUENCES: 85  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP  
STREET: 1201 New York Avenue, N.W., Suite 1000  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
Zip: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/10220  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/409,305  
FILING DATE: 24-MAR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/348,824

FILING DATE: 29-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08-308,104  
FILING DATE: 16-SEP-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/300,266  
FILING DATE: 02-SEP-1994  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/289,221  
FILING DATE: 12-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Ihnen, Jeffrey L.  
REGISTRATION NUMBER: 28,957  
REFERENCE/DOCKET NUMBER: 24884-109347  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-962-4810  
TELEFAX: 202-962-8300  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6769 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
PCT-US95-10220-20

Query Match 15.5%; Score 226.2; DB 5; Length 6769;  
Best Local Similarity 87.0%; Pred. No. 9.5e-48;

Matches 260; Conservative 0; Mismatches 38; Indels 1; Gaps 1;  
QY 1162 GCTGGGAGGGTGGCTCCTGTAATCCAGCAGCTTTGGAGGC-GAGGAGCAGAGAT 1220  
|||||  
DB 5743 GCTGGGAGGGTGGCTCCTGTAATCCAGCAGCTTTGGAGGCAGGAGGAGG 5684  
QY 1221 CACTTGAGTGAAGATTCAGACACCCCTGGCCAAATGTAATCTCTCTCTCTA 1280  
|||||  
DB 5683 CACTTGAGTGAAGATTCAGACACCCCTGGCCAAATGTAATCTCTCTCTA 5624  
QY 1281 AAAATACAAAGTATGCTGATATGTCGAGGCGCTGTATCCAGCTACTTGGAGG 1340  
|||||  
DB 5623 AAAATACAAAGTATGCTGATATGTCGAGGCGCTGTATCCAGCTACTTGGAGG 5564  
QY 1341 CTGAGGAGGAGATTCCTGAATATGAGGAGGAGGAGTTGAGTTGAGATCAGAC 1400  
|||||  
DB 5563 CTGAGGAGGAGATTCCTGAATATGAGGAGGAGGAGTTGAGTTGAGATCAGAC 5504  
QY 1401 CACTATCTCCAGCTGGGCGACAGAGTAACTCTCTCAAAAAAAAAAAAAAAAAA 1459  
|||||  
DB 5503 CTCTGCTCTCAGCTGGGCGACAGAGTAACTCTCTCAAAAAAAAAAAAAAAAAA 5445

## RESULT 10

US-08-451-777A-7/c  
Sequence 7, Application US/08451777A  
Patent No. 5789223  
GENERAL INFORMATION:  
APPLICANT: Bergsma, Derek J.  
APPLICANT: Stambolian, Dwight  
TITLE OF INVENTION: Human Galactokinase Gene  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
STREET: 709 Swedeland Road/UM2220  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406-0939

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/451,774  
FILING DATE: 26-MAY-1995  
CLASSIFICATION: 436  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/10825  
FILING DATE: 23-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Eagle, Ailsa M.  
REGISTRATION NUMBER: 37,126  
REFERENCE/DOCKET NUMBER: P50268-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-5364  
TELEFAX: 610-270-5090  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7676 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-451-777A-7

Query Match 15.0%; Score 219.2; DB 3; Length 7676;  
Best Local Similarity 85.3%; Pred. No. 5.7e-46;  
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;

QY 1161 GCTGGGAGGGTGGCTCCTGTAATCCAGCAGCTTTGGAGGC-GAGGAGGAG 1219  
|||||  
DB 6058 GGTGGGAGGGTGGCTCCTGTAATCCAGCAGCTTTGGAGGCAGGAGGAG 5999  
QY 1220 TCAGTTGAGTGAAGATTCAGACACCCCTGGCCAAATGTAATCTCTCTCTACT 1279  
|||||  
DB 5998 TTGCTGATGATCAGAGCTGAGACCCCTGGCCAAATGTAATCTCTCTCTACT 5939  
QY 1280 AAAATACAAAGTATGCTGATATGTCGAGGCGCTGTATCCAGCTACTTGGAGG 1339  
|||||  
DB 5938 AAAATACAAAGTATGCTGATATGTCGAGGCGCTGTATCCAGCTACTTGGAGG 5879  
QY 1340 GCTGAGGAGGAGATTCCTGAATATGAGGAGGAGGAGTTGAGTTGAGATCACA 1399  
|||||  
DB 5878 GCTGAGGAGGAGATTCCTGAATATGAGGAGGAGGAGTTGAGTTGAGATCAG 5819  
QY 1400 CCACTATCTCCAGCTGGGCGACAGAGTAACTCTCTCAAAAAAAAAAAAAAAAAA 1459  
|||||  
DB 5818 CCACTATCTCCAGCTGGGCGACAGAGTAACTCTCTCAAAAAAAAAAAAAAAAAA 5759

## RESULT 11

US-08-451-778A-7/c  
Sequence 7, Application US/08451778A  
Patent No. 5830649  
GENERAL INFORMATION:  
APPLICANT: Bergsma, Derek J.  
APPLICANT: Stambolian, Dwight  
TITLE OF INVENTION: Human Galactokinase Gene  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
STREET: 709 Swedeland Road/UM2220  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

```
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,778A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 800
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/10825
FILING DATE: 23-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Eagle, Alissa M.
REGISTRATION NUMBER: 37,126
REFERENCE/DOCKET NUMBER: P50268-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5364
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7676 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-451-778A-7
```

```
Query Match      15.0%: Score 219.2; DB 4; Length 7676;
Best Local Similarity 85.3%: Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
```

```
QY 1161 GCGTGGGAGGGTGGCTGATGCTGTATCCAGCACTTGGGAGGC-GAGGCAGGCGAGA 1219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6058 GGGTGGGCGCGGTGGCTGACACCTGTATCCAGCACTTGGGAAGCTGAGGAGGAGCA 5999

QY 1220 TCACCTTGAGGTGAAGAGGTTCAAGACCGCTGGCCAAAATGGTGAATCTGTCTACT 1279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5998 TGGCTTGAGGTGAAGAGGTTCAAGACCGCTGGCCAAAATGGTGAATCTGTCTACT 5939

QY 1280 AAAAATACAAAAGTTAGCTGATATGATGAGGAGGCGCTGTATCCAGCTACTTGGGAG 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5938 AAAAATACAAAAGTTAGCTGATATGATGAGGAGGCGCTGTATCCAGCTACTTGGGAG 5879

QY 1340 GCTGAGGAGGAGAAATGCTTGAATATGGAGGAGCAGAGGTTGAATGAGTTGAGATCACA 1399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5878 GCTGAGGAGGAGAAATGCTTGAATATGGAGGAGGAGGTTGAGATGAGATCAGCANG 5819

QY 1400 CCACCTACTCCAGCTGGGGCAGACAGATAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5818 CCACCTGCACTTCCAGCTGGGGTGAACAGATGAGGCTCCCTCTCAAAAAAAAAAAAAAGA 5759
```

RESULT 12  
US-08-998-208-7/c

Sequence 7, Application US/08998208  
Patent No. 5880105  
GENERAL INFORMATION:  
APPLICANT: Bergsma, Derk J.  
APPLICANT: Stambolian, Dwight  
TITLE OF INVENTION: Human Galactokinase Gene  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESSEE: Intellectual Property  
STREET: 709 Swedeland Road/W22220  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,208
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/451,777
FILING DATE: 26-MAY-1995
APPLICATION NUMBER: PCT/US94/10825
FILING DATE: 23-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Eagle, Alissa M.
REGISTRATION NUMBER: 37,126
REFERENCE/DOCKET NUMBER: P50268-1B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5364
TELEFAX: 610-270-5090
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 7676 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-998-208-7
```

```
Query Match      15.0%: Score 219.2; DB 4; Length 7676;
Best Local Similarity 85.3%: Pred. No. 5.7e-46;
Matches 256; Conservative 0; Mismatches 43; Indels 1; Gaps 1;
```

```
QY 1161 GCGTGGGAGGGTGGCTGATGCTGTATCCAGCACTTGGGAGGC-GAGGCAGGCGAGA 1219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 6058 GGGTGGGCGCGGTGGCTGACACCTGTATCCAGCACTTGGGAAGCTGAGGAGGAGCA 5999

QY 1220 TCACCTTGAGGTGAAGAGGTTCAAGACCGCTGGCCAAAATGGTGAATCTGTCTACT 1279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5998 TGGCTTGAGGTGAAGAGGTTCAAGACCGCTGGCCAAAATGGTGAATCTGTCTACT 5939

QY 1280 AAAAATACAAAAGTTAGCTGATATGATGAGGAGGCGCTGTATCCAGCTACTTGGGAG 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5938 AAAAATACAAAAGTTAGCTGATATGATGAGGAGGCGCTGTATCCAGCTACTTGGGAG 5879

QY 1340 GCTGAGGAGGAGAAATGCTTGAATATGGAGGAGCAGAGGTTGAATGAGTTGAGATCACA 1399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5878 GCTGAGGAGGAGAAATGCTTGAATATGGAGGAGGAGGTTGAGATGAGATCAGCANG 5819

QY 1400 CCACCTACTCCAGCTGGGGCAGACAGATAGACTGTCTCAAAAAAAAAAAAAAAAAAAAA 1459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 5818 CCACCTGCACTTCCAGCTGGGGTGAACAGATGAGGCTCCCTCTCAAAAAAAAAAAAAAGA 5759
```

RESULT 13  
PCT-US95-06743-7/c

Sequence 7, Application PC/TUS9506743  
GENERAL INFORMATION:  
APPLICANT: Bergsma, Derk J.  
APPLICANT: Stambolian, Dwight  
TITLE OF INVENTION: Human Galactokinase Gene  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corp./Corporate  
ADDRESSEE: Intellectual Property  
STREET: 709 Swedeland Road/W22220  
CITY: King of Prussia  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

Query Match	14.6%;	Score 213.4;	DB 1;	Length 26764;
Best Local Similarity	80.8%;	Pred. No. 2.7e-44;		
Matches 261; Conservative	0;	Mismatches 61;	Indels 1;	Gaps 1;

[illegible]

RESULT 15  
 US-08-370-975B-6  
 Sequence 6, Application US/08370975B  
 Patent No. 5622851  
 GENERAL INFORMATION:  
 APPLICANT: Maley, Frank  
 APPLICANT: Maley, Gladys F.  
 APPLICANT: Melner, Karen X.B.  
 TITLE OF INVENTION: Human deoxycytidylate deaminase gene  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
 STREET: Clinton Square, P.O. Box 1051  
 CITY: Rochester  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 14603  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/370,975B  
FILING DATE: 10-JAN-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Timian, Susan J.  
REGISTRATION NUMBER: 34,103  
REFERENCE/DOCKET NUMBER: 20894/80  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716)263-1636  
TELEFAX: (716)263-1600  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20303 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT: 4q35  
US-08-370-975B-6

Query Match 14.6%; Score 213.4; DB 1; Length 20303;  
Best Local Similarity 80.8%; Pred. No. 2,4e-44;  
Matches 261; Conservative 0; Mismatches 61; Indels 1; Gaps 1;  
QY 1137 GGAATAATCAAGACTCTACAAAGAGGCTGGGAGGCTCATGCTGTATCCAGCA 1196  
DB 592 GGTACAGCCAAAGCCAGAGTGAAGGTGGGACGATGCTCAGCCTGTATCCAGCA 651  
QY 1197 CTTTGGAGG-CGAGGAGGAGATCACTTGAGGTAAAGATTCAAGACCAGCCTGGCCA 1255  
DB 652 CTTTGGAGGTGGAGGAGGAGATCACTGAGGTTCGAGATCCAGACCAGCCTAGCCA 711  
QY 1256 AATGTGAATCTCTCTACTATAAAATACAAAAGTGTGATATGTTGGAGGCG 1315  
DB 712 ACATGGTGAACCCCATCTCTACTATAAAATACAAAATTAGCTGTGCTGGTGGAGGTG 771  
QY 1316 CCGTATATCCAGCTACTTGGAGGCTGAGGAGAGAAATTGCTTGAATATGGAGGAG 1375  
DB 772 CCGTATATCTCAGCTACTCTGGAGGCTGAGGAGAAAGAAATCGCTTGAACTCAGAGGTG 831  
QY 1376 AGGTGAAGTGAAGTGAAGATCAGACCACCTATCTCCAGCTGGGGCAACAGAGTAAGATC 1435  
DB 832 AGGTGTAGTGAAGCAAGACTGCGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTC 891  
QY 1436 TGTCTCAAAAAAAAAAAAAA 1458  
DB 892 TGTCTCAAAAAAAAAAAAAA 914

Search completed: September 28, 1999, 11:33:47  
Job time: 2107 sec

```

      :::: ||||| ::::|::| ::::|::|
      236 rsnalatrpilealaglnthrilealalaasn 247

```

```

seq_name: sp Vertebrate:042160

```

```

seq_documentation_block:

```

```

ID_042160: PRELIMINARY; PRT; 245 AA.

```

```

AC_042160:

```

```

DT 01-JAN-1998 (TREMBlrel. 05, Created)

```

```

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

```

```

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

```

```

DE TRIPSINOGEN B2 PRECURSOR (FRAGMENT).

```

```

GN TRYPB2.

```

```

OS Petromyzon marinus (Sea lamprey).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;

```

```

OC Petromyzontiformes; Petromyzontidae; Petromyzon.

```

```

RN [1]

```

```

RP SEQUENCE FROM N.A.

```

```

RA ROACH J.C.;

```

```

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL: AF011901; AAB69657.1; -.

```

```

DR PRAM: PF00089; trypsin; 1.

```

```

KW Signal.

```

```

FT SIGNAL 1 1

```

```

FT NON_TER 1 1

```

```

FT SIGNAL <1 13 POTENTIAL.

```

```

FT CHAIN 14 245 TRYPsin B2.

```

```

SQ SEQUENCE 245 AA; 26001 MW; 071872F0 CRC32;

```

```

alignment_scores:

```

```

Quality: 423.50 Length: 245

```

```

Ratio: 2.664 Gaps: 6

```

```

Percent Similarity: 64.898 Percent Identity: 36.327

```

```

alignment_block:

```

```

US-09-030-606-173 x 042160 . .

```

```

Align seg 1/1 to: 042160 from: 1 to: 245

```

```

10 CACTGGCAGCCCTGGCAGGCGCAGTCATGAAACGAATTGTTCTG 59

```

```

31 HlsserGlnProTlp6lnValSerLeuAnllleGlyTyrHis...PheCy 46

```

```

60 CTCGGGCGCTGTGTCATCCGAGTGGGTGCTGTCAGCCGACACTGT 109

```

```

46 s6llylSerLeuIleSerSerGlnTlp6lnValSerAlaAlaHisCysT 63

```

```

110 TCACGAACTCC.....TACACCATCGGGCTGGGCTGCACAGCTTTGAG 153

```

```

63 yrglnThrAlaSerArgIleSerValArgIleGlyLuhHisnIlePhe 79

```

```

154 GCCGACCAAGAGCCAGGAGGAGCATGTGAGGCCAGCCTCTCCGTAGC 203

```

```

80 Val...ThrglGlyThrGlnGlnArgIleGlnAlaSerLysAlaIleAr 95

```

```

204 GCACCCAGAGTACACAGACACCTTGTCTAAGACCTCATGCTCATCA 253

```

```

95 ghIsPrgInTyrAsnSerAlaThrIleAspAsnspIleMetLeuIleL 112

```

```

254 AGTTGACGAATCCGTGTGCGAGTGTGACACCATCCGAGCATCAGATT 303

```

```

112 yelSerSerProAlaThrLeuAsnGlnTyrAlaGlnAlaIleProLeu 128

```

```

304 GCTTGCACTGCCCTACCGGGGAACTTGGCTCTGTTCTGGCTGGGG 353

```

```

129 ProSerSerCyValGlyThrGlyValMetCysThrIleSerGlyTyrpI 145

```

```

354 TCTGCTGGGGAACGTCAGCTCAGCGGTGTGTCTGCCCTCTCAAGGA 403

```

```

145 Y.....

```

```

404 GGTCTCTGCCAGTCGCGGGGCTGACCCAGAGCTCTGCTCCAGGCA 453

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

.....

```

```

146 .....GluThrGlnThrSerValGly 152
454 GAATGCTACCGTGTGCTGCGAGTGGTGAACGAGTGTGCTGAGGAG 503
153 Ser..ProAspValLeuMetCysValGlnAlaProValLeuSerAspThr 168
504 GTCTGCACTAAGCTCTATGACCCGCTGTACACCCACCATGTTCTGCGC 553
159 SerCysArgAsnSerTyr-ProGlyAspIleThrAsnAsnMetIleCysLe 185
554 CGCGGAGGCGAAGACCAAGAGGACTCTGCAACGCTGCTGGGGGCGC 603
185 uGlyTyrLeuGlnGlyGlyLysAspSerCysGlnIleLysAspSerGlyGlyP 202
604 CCCGTATGTGCAAGCGTACTGTCAGGCGCTTGTCTTGGGAAAAGCC 653
202 rovalValCysAsnGlyLneugInglyIleValSerTlp6lnYargGly 218
654 CCGTGTGCGCAAGTGGCGTGCAGTGTCTACACCAACCTGCAATT 703
219 ..CysAlaLeuProAsnTyr-ProGlyValTyrThrLysValCysAsnTyr 234
704 CACTGAGTGGATGAGAAAACCGTCCAGCGCAGT 737
234 rAsnSerTlp6lnAlaSerThrMetAlaAlaasn 245

```

---



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:29:30 ; Search time 1809.22 Seconds  
(without alignments)  
1379.188 Million cell updates/sec

Title: US-09-030-606-173  
Perfect score: 1265  
Sequence: 1 GGCAGCCCGCCACTCCGAGCC.....AGAGANGCAGCAAAAAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 98626752 residues

Database :

EST: \*  
1: em\_est1: \*  
2: em\_est2: \*  
3: em\_est3: \*  
4: em\_est4: \*  
5: em\_est5: \*  
6: em\_est6: \*  
7: em\_est7: \*  
8: em\_est8: \*  
9: em\_est9: \*  
10: em\_est10: \*  
11: em\_est11: \*  
12: em\_est12: \*  
13: em\_est13: \*  
14: em\_est14: \*  
15: em\_est15: \*  
16: em\_est16: \*  
17: em\_est17: \*  
18: em\_est18: \*  
19: em\_est19: \*  
20: em\_est20: \*  
21: em\_est21: \*  
22: em\_est22: \*  
23: em\_est23: \*  
24: em\_est24: \*  
25: em\_est25: \*  
26: em\_est26: \*  
27: em\_est27: \*  
28: em\_est28: \*  
29: em\_est29: \*  
30: em\_est30: \*  
31: em\_est31: \*  
32: em\_est32: \*  
33: em\_est33: \*  
34: em\_est34: \*  
35: em\_est35: \*  
36: em\_est36: \*  
37: em\_est37: \*  
38: em\_est38: \*  
39: em\_est39: \*  
40: em\_est40: \*  
41: em\_est41: \*  
42: em\_est42: \*  
43: em\_est43: \*  
44: em\_est44: \*  
45: em\_est45: \*  
46: em\_est46: \*  
47: em\_est47: \*  
48: em\_est48: \*  
49: em\_est49: \*  
50: em\_est50: \*  
51: em\_est51: \*  
52: em\_est52: \*  
53: em\_est53: \*

54: em\_est22: \*  
55: em\_est23: \*  
56: em\_est24: \*  
57: em\_est25: \*  
58: em\_est26: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	476.8	37.7	777	48	A1557281	A1557281 PT2.1_15_
2	355.2	28.1	404	34	AA503963	AA503963 nh39a01.s
3	315.4	24.9	569	50	A1686689	A1686689 tu35g11.x
4	305.6	24.2	415	35	AA551449	AA551449 nj55e05.s
5	290.2	22.9	576	34	AA533140	AA533140 nj46b06.s
6	241.2	19.1	722	48	A1557025	A1557025 PT2.1_10_
7	209.2	16.5	259	35	AA565161	AA565161 nk32c07.s
8	199.8	15.8	229	35	AA552779	AA552779 nk57a10.s
9	192.8	15.2	241	36	AA603977	AA603977 no47a09.s
10	192	15.2	218	50	A1674521	A1674521 wc39g02.x
11	176	13.9	216	32	AA336074	AA336074 EST40866
12	167.2	13.2	188	50	A1675815	A1675815 wb97b12.x
13	155	12.3	191	35	AA595489	AA595489 no34h04.s
14	115.8	9.2	507	35	AA583052	AA583052 nm80g09.s
15	109.8	8.7	517	31	AA293027	AA293027 zt54a12.x
16	108.4	8.6	498	26	W73168	W73168 zds5e11.r1
17	108.2	8.6	505	45	A1391329	A1391329 mb71b07.y
18	101.2	8.0	590	33	AA411352	AA411352 zc33b03.r
19	101	8.0	321	31	AA302930	AA302930 EST113097
20	99.6	7.9	333	49	A1636241	A1636241 tz93c03.x
21	98.8	7.8	686	46	A1415008	A1415008 mb71h07.x
22	97.4	7.7	586	26	W73140	W73140 zds5e11.s1
23	96	7.6	356	39	AA838788	AA838788 a187g08.s
24	95.6	7.6	504	49	A1620091	A1620091 ty48d10.x
25	93.2	7.4	212	28	C16642	C16642 C16642 Clon
26	91.4	7.2	420	43	A1226226	A1226226 ue88f06.y
27	90.8	7.2	367	41	A1002101	A1002101 ct38d03.s
28	90	7.1	400	31	AA293231	AA293231 zt56g09.r
29	90	7.1	539	34	AA477689	AA477689 zu44a12.r
30	89.8	7.1	480	39	AA846771	AA846771 a141f01.s
31	89	7.0	234	33	AA452459	AA452459 zx29g09.r
32	86.8	6.9	519	33	AA401397	AA401397 zu68b01.s
33	86	6.8	122	49	A1632842	A1632842 tz32h08.x
34	85.4	6.8	396	49	AV005299	AV005299 AV005299
35	85.2	6.7	493	45	A1324874	A1324874 m185b05.x
36	85.2	6.7	363	47	A1477594	A1477594 fb57g08.x
37	84.4	6.7	392	25	W16362	W16362 mb56d06.r1
38	84.4	6.7	457	45	A1385433	A1385433 m185b05.y
39	84.2	6.7	1029	25	W13212	W13212 ma83a07.r1
40	84	6.6	499	43	A1178581	A1178581 EST322260
41	83.8	6.6	626	36	AA643312	AA643312 nr59b08.s
42	83.6	6.6	456	42	A1139437	A1139437 qc20e03.x
43	83.2	6.6	585	35	C23111	C23111 C23111 Japa
44	82.8	6.5	599	43	A1237604	A1237604 EST34166
45	82.8	6.5	440	46	A1411685	A1411685 EST339979

#### ALIGNMENTS

RESULT 1  
LOCUS A1557281 777 bp mRNA  
DEFINITION PT2.1\_15\_G12.r tumor2 Homo sapiens cDNA 3', mRNA sequence.  
ACCESSION A1557281  
NID 94489644  
VERSION A1557281.1 GI:4489644  
KEYWORDS EST.

SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
AUTHORS	1 (bases 1 to 777) Huang,G.M., Ng,W., Farkas,J., Chen,L., Liang,H.A., Gordon,D., Jun Yu,J. and Hood,L.
TITLE	Prostate Cancer Expression Profiling by cDNA Sequencing Analysis
JOURNAL	Unpublished (1999)
COMMENT	On May 16, 1998 this sequence version replaced gi:3138767.

**CONTACT:** Guyang Matthew Huang  
 Leroy Hood  
 University of Washington  
 Department of Molecular Biotechnology, Box 357730, University of  
 Washington, Seattle, WA 98195  
 Tel: 5106280100  
 Fax: 5106280108  
 Email: huanggm@yahoo.com,  
 Location/Qualifiers  
 I. .777

BASE COUNT	162 a	205 c	227 g	157 t	26 others
ORIGIN					

Query Match	37.7%	Score 476.8;	DB 48;	Length 777;
Best Local Similarity	81.9%	Pred. No. 9.6e-106;		
Matches 643;	Conservative	0;	Mismatches 48;	Indels 94;
				Gaps 5

OY	68	TCCTGATGCATCCGACAGGGGTGCTGTACGCCGACACTGTTTCCAAACATCCGACCA	127
Db	75	TCCTGTGCATCCGACAGGGGTGCTGTACGCCGACACTGTTTCCAAACATCTTACACCA	134
OY	128	TCGGGCTGGGCTTCACAGTCTTGAGGCGCACCAAGACCAGGAGCCAGATGGTGAGG	187
Db	135	TCGGGCTGGGCTTCACAGTCTTGAGGCGCACCAAGACCAGGAGCCAGATGGTGAGG	194
OY	188	CCAGGCTTCGCTGAGGCGACCCAGATGACAACACCTTGTGCTGAAGACTCATGC	247
Db	195	CCAGGCTTCGCTGAGGCGACCCAGATGACAACACCTTGTGCTGAAGACTCATGC	254
OY	248	TCATCAAGTTGGAGCAATCCGTGTCCGAGTGTGACACCATCCGAGCATCAGATTGCTT	307
Db	255	TCATCAAGTTGGAGCAATCCGTGTCCGAGTGTGACACCATCCGAGCATCAGATTGCTT	314
OY	308	CGCAGTGCCTTACCCGCGGGGAACCTTGTGCTCTGTTTGGCTGGGGTCTGCTGGCAACG	367
Db	315	CGCAGTGCCTTACCCGCGGGGAACCTTGTGCTCTGTTTGGCTGGGGTCTGCTGGCAACG	374
OY	368	GTGAGCTACAGGGGTGTGTTGTGCTTCCCTTTCAGAGAGGTCTCTGCCCCAGTCGGGGGGC	427
Db	375	G-----	376
OY	428	TGACCCAGAGCTCTGCGTCCAGGACAGAAATCCCTACCGTGTGCAAGTGCSTGAACGTCT	487
Db	376	-----CAGAAAGTACCGTGTGTCAGTGCAGTGCAGAACGTCT	410
OY	488	GGTGTGCTCTAGAGAGGTCTGTCAATAAGCTCTATAGACCCGCTGTACCAACCCACAGATGT	547
Db	411	GGTGTGCTCTAGAGAGGTCTGTCAATAAGCTCTATAGACCCGCTGTACCAACCCACAGATGT	470

Accession	Sequence	Position
Oy	CTGGCCCGCGAGGGCAAGACCAAGAGAACTCTTGCACAGGTACTCTGGGGGGCCCT	607
Oy	548 CTGGCCCGCGAGGGCAAGACCAAGAGAACTCTTGCACAGGTACTCTGGGGGGCCCT	607
Db	471 CTGGCCCGCGAGGGCAAGACCAAGAGAACTCTTGCACAGGTACTCTGGGGGG-CCTT	529
Oy	608 GATCTGCAACGGGTACTTGCAGGGCCCTGTGTCTTTTGGAAAAAGCCCGTGTGGCCAA	657
Db	530 GATTGTGAAACCGGGGNCCTTGTGACGGCCCTTGTGTCTTTGGAAAAAGCC--GTGTGGCAAGT	587
Oy	668 TGGCGTGGCAGGTGTCTACACCAACCTCTGCAATTTCACAGGTGGATACAGAAACCGT	727
Db	568 TGGCGTGGCAGGTGTCTACCAAC-----CTGCAATTACTAGAGGGATACANAAAAACGGT	642
Oy	728 CCAGGCCAGTTAACTCTGGGGACTTGGGAACCATGAATATGACCCCAATATACATCTGC	787
Db	643 C--AGCGAGTTAACTTTGGGGAGTGGANNCATTAATATGNCGCCCAATAATATCTGGG	700
Oy	788 GGAAG 792	
Db	701 GAAG 705	

RESULT	2
AA503963/c	
LOCUS	404 bp mRNA
DEFINITION	h339601.s1 NCI CGAP pr Homo sapiens cDNA clone IMAGE:554656 similar to SW:TLK_MOUSE P15946 GIANTULAR KALIKREIN K11 PRECURSOR containing MSRI.b2 MSRI repetitive element ; , mRNA sequence.
ACCESSION	AA503963
NID	92238930
VERSION	AA503963.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
<i>Homo sapiens</i>	Elkaytor; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	1 (bases 1 to 404)	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).	
			Tumor Gene Index	Unpublished (1997)	On May 8, 1995 this sequence version replaced gi:801235.

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: David G. Bestwick, M.D., Rodrigo F. Chuagut,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Kitzman, Ph.D.  
cDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
[www-bio.liml.gov/db/rr/image/image.html](http://www-bio.liml.gov/db/rr/image/image.html)

Insert Length: 1244    Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 144.

**FEATURES**  
**Source**

```

source
1. .404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:954696"
/clone_id="NCI_CGAP_Pr5"
/sex="male"
/tissue_type="prostate"
/lab_host="DH10B"
/note="Vector: pAMP10: mRNA made from normal prostatic
epithelial cells, cDNA made by oligo-dT priming.
Non-directionally cloned. Size-selected on agarose gel
average insert size 600 bp."
BASE COUNT      64 a      159 g      90 t
ORIGIN

```



REFERENCE 1 (bases 1 to 415)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On May 9, 1995 this sequence version replaced g1:802243.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
cDNA Library Preparation: David B. Krizman, Ph.D.  
DNA Sequencing by: Greg Lennon, Ph.D.  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

Insert Length: 640 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from AmerSham  
High quality sequence stop: 412.

FEATURES  
source Location/Qualifiers

1..415  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:96416"  
/clone\_lib="NCI\_CGAP\_P19"  
/sex="male"  
/tissue\_type="normal prostatic epithelial cells"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: PAMP10; mRNA made from  
normal prostatic epithelial cells, cDNA made by oligo-dT  
priming. Non-directionally cloned. Size-selected on  
agarose gel, average insert size 600 bp. Library made by  
D. Krizman, NIH."

BASE COUNT 78 a 125 c 129 g 83 t

ORIGIN

Query Match 24.2%; Score 305.6; DB 35; Length 415;

Best Local Similarity 81.3%; Pred. No. 2,2e-64; Mismatches 9; Indels 84; Gaps 1;

Matches 405; Conservative 0; Mismatches 9; Indels 84; Gaps 1;

149 TTGAGCCGACGACAGAGCCGAGGAGCCAGATGGTGGAGCCACCTCTCCGTCAGGACCG 208

2 TCGACGCCGACGACAGAGCCGAGGAGCCAGATGGTGGAGCCACCTCTCCGTCAGGACCG 61

209 CAGAGTACACAGACAGCCCTGCTGCTGCTACAGACCTCATGCTCATCAAGTTGGACGAATCG 268

62 CAGAGTACACAGACAGCCCTGCTGCTGCTACAGACCTCATGCTCATCAAGTTGGACGAATCG 121

269 TGTCCAGTGTGACACCATTCGAGCATGACATGCTTGTGCGAGTCCCTACCGCGGGA 328

122 TGTCCAGTGTGACACCATTCGAGCATGACATGCTTGTGCGAGTCCCTACCGCGGGA 181

329 ACTCTTGCCCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 388

182 ACTCTTGCCCTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 222

389 TGCCCTCTTCAGAGAGGTCTCTGCCCCAGTCCGCGGGGCTGACCCAGAGCTCTGCTCCC 448

222 ----- 222

449 AGGCAAAATGCTTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 508

222 ----CAGATGCTTACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 277

509 CAGTAACTTATGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 568

278 CAGTAACTTATGACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 337

569 GCAGAAAGACTCTCTGACAGCGTACTCTGCGGGGCGCCCTGATCTGACAGCGGTAATTGCA 628

Db 338 CAGAGGAGCTCTCTGACAGCGTACTCTGCGGGGCGCCCTGATCTGCAACGGGTAATTGCA 397

QY 629 GGGCCTTGCTCTTCGG 646

Db 398 GGGCCTTGCTCTTCGG 415

RESULT 5

LOCUS AA533140

DEFINITION AA533140 576 bp mRNA EST 21-AUG-1997

ACCESION n146h06.s1 NCI\_CGAP\_P19 Homo sapiens cDNA clone IMAGE:995579

VERSION Similar to TR:G198491 G198491 KALLIREIN: contains MSRI.tl MSRI

KEYWORDS repetitive element ;, mRNA sequence.

ORGANISM human.

REFERENCE Homo sapiens

AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

TITLE Eutheria; Primates; Catarrhini; Hominiidae; Homo.

JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced g1:1406821.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquai,

M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

[www-bio.llnl.gov/dbfp/image/image.html](http://www-bio.llnl.gov/dbfp/image/image.html)

Insert Length: 885 Std Error: 0.00

Seq primer: -40m13 fwd. ET from AmerSham

High quality sequence stop: 256.

FEATURES

source Location/Qualifiers

1..576  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:995579"  
/clone\_lib="NCI\_CGAP\_P19"  
/sex="male"  
/tissue\_type="normal prostatic epithelial cells"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: PAMP10; mRNA made from  
normal prostatic epithelial cells, cDNA made by oligo-dT  
priming. Non-directionally cloned. Size-selected on  
agarose gel, average insert size 600 bp. Library made by  
D. Krizman, NIH."

BASE COUNT 169 a 156 c 139 g 102 t 10 others

ORIGIN

Query Match 22.9%; Score 290.2; DB 34; Length 576;

Best Local Similarity 95.7%; Pred. No. 1.3e-60; Mismatches 13; Indels 1; Gaps 1;

Matches 309; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

567 GACCAAGAGACTCTCTGACAGCGTACTCTGCGGGGCGCCCTGATCTGCAACGGGTAATT 625

42 GACCAAGAGACTCTCTGACAGCGTACTCTGCGGGGCGCCCTGATCTGCAACGGGTAATT 101

626 GCAGGGCCTTGCTGCTTTCGGAAGAGCCCGGTGGCCAGTTGGCGTCCAGGTGCTTA 685

102 GCAGGGCCTTGCTGCTTTCGGAAGAGCCCGGTGGCCAGTTGGCGTCCAGGTGCTTA 161

QY 686 CACCAACCTGTGCAAAATTCAGTAGAGAAAACGTCAGGACCACTACTCTG 745  
 |||||||  
 Db 162 CACCAACCTGTGCAAAATTCAGTAGAGAAAACGTCAGGACCACTACTCTG 221  
 |||||||  
 QY 746 GGGACTGGGAACCCATGAAATGACCCCAATATCATCTGCGGAGGAATTCAGATA 805  
 |||||||  
 Db 222 GGGACTGGGAACCCATGAAATGACCCCAATATCATCTGCGGAGGAATTCAGATA 281  
 |||||||  
 QY 806 TCTGTTCCCAAGCCCTCTCTCTCTCAGGCCAGAGAGTCAGGCCCCCCTCTCTCT 865  
 |||||||  
 Db 282 TCTGTTCCCAAGCCCTCTCTCTCTCAGGCCAGAGAGTCAGTCCCAAGAACTACTTCT 341  
 |||||||  
 QY 866 CAACCAAGGATACAGATCCCA 888  
 |||||  
 Db 342 TCAACCAAGGATACAGATCCCA 364  
 |||||

RESULT 6  
 LOCUS A1557025 722 bp mRNA EST 23-MAR-1999  
 DEFINITION PT2.1\_10.F05.r tumor2 Homo sapiens cDNA 3', mRNA sequence.  
 ACCESSION A1557025  
 NID 94489388  
 VERSION A1557025.1 GI:4489388  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 722)  
 AUTHORS Huang, G.-M., Ng, W., Farakas, J., Chen, L., Liang, H.A., Gordon, D., Jun  
 Yu, J. and Hood, L.  
 TITLE Prostate Cancer Expression Profiling by cDNA Sequencing Analysis  
 JOURNAL Unpublished (1999)  
 COMMENT On May 18, 1998 this sequence version replaced gi:3138511.

CONTACT: Guyan Matthew Huang  
 Leroy Hood  
 University of Washington  
 Department of Molecular Biotechnology, Box 357730, University of  
 Washington, Seattle, WA 98195  
 Tel: 5106280100  
 Fax: 5106280108  
 Email: huangm@u.washington.edu  
 Location/Qualifiers  
 1..722  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="15:21q"  
 /clone\_11b="tumor2"  
 /note="Organ: Prostate; Vector: pBluescript; Directional  
 cDNA library was constructed using Lambda ZP II kit  
 (Stratagene). mRNA was extracted from a frozen prostate  
 tumor tissue (Mayo Clinics)."  
 BASE COUNT 146 a 156 c 195 g 176 t 49 others  
 ORIGIN

Query Match 19.1%; Score 241.2; DB 48; Length 722;  
 Best Local Similarity 89.7%; Pred. No. 1.1e-48;  
 Matches 297; Conservative 0; Mismatches 26; Indels 8; Gaps 4;  
 QY 115 AACTCTACACCATGGGCTGGGCTGACAGTTTGAGGCCGACCAAGAGCCAGGGAGC 174  
 |||||||  
 Db 115 AGCTTCTACACCATGGGCTGGGCTGACAGTTTGAGGCCGACCAAGAGCCAGGGAGC 174  
 |||||||  
 QY 175 CAGATGGGAGGCGCAGCCTCTCGTACGGCAGCAGGATCAACAGACCTCTGCTGCT 234  
 |||||||  
 Db 175 CAGATGGGAGGCGCAGCCTCTCGTACGGCAGCAGGATCAACAGACCTCTGCTGCT 234  
 |||||||  
 QY 235 AACGACCTCATCTCATCAAGTTGGAGCAATCGTGTCCAGTCTGACACCATCCGGAGC 294  
 |||||||

Db 235 AACGACCTCATCTCATCAAGTTGGAGCAATCGTGTCCAGTCTGACACCATCCGGAGC 294  
 |||||||  
 QY 295 ATGAGCATGCTCTGCGAGCTGCTACCGGGGGAACCTTGCC-TGTTCTGGCTGGGG 353  
 |||||||  
 Db 295 ATGAGCATGCTCTGCGAGCTGCTACCGGGGGAACCTTGCTTCTGTTCTGGCTGGGG 354  
 |||||||  
 QY 354 TCTGC-TGGCGAAGCGGTACGAGGTGTGTGTGCTGCTCTTC---AAGAGAGTCC 408  
 |||||||  
 Db 355 TCTGCTTGGGGAACGGTACGAGGTGTGTGTGCTGCTCTTC---AAGAGAGTCC 414  
 |||||||  
 QY 409 TCTGCCCA--GTGCGGGGGCTGACCCAGAG 437  
 |||||  
 Db 415 TTTTNCNAAGTNTGCGGGGCTACNCCAGAG 445  
 |||||

RESULT 7  
 LOCUS AA565161/c 259 bp mRNA EST 08-SEP-1997  
 DEFINITION nk52c07.s1 NCI-CGAP\_P7 Homo sapiens cDNA clone IMAGE:1017132  
 similar to contains MSRL.b3 MSRL MSRL repetitive element ;, mRNA  
 sequence.  
 ACCESSION AA565161  
 NID 92336800  
 VERSION AA565161.1 GI:2336800  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 259)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1397569.

CONTACT: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: David G. Bostwick, M.D., Rodrigo F. Chuquil,  
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: David B. Krizman, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www-bio.llnl.gov/dbp/image/image.html

Insert Length: 458 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 258.  
 Location/Qualifiers  
 1..259  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="17q21"  
 /clone\_11b="NCI-CGAP\_P7"  
 /sex="male"  
 /tissue\_type="Low-grade prostatic neoplasia"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: PAMP10; mRNA made from  
 prostate intraepithelial neoplasia (high-grade), cDNA  
 made by oligo-dT priming. Non-directionally cloned.  
 Size selected on agarose gel, average insert size 600 bp."

BASE COUNT 56 a 41 c 106 g 56 t  
 ORIGIN  
 Query Match 16.5%; Score 209.2; DB 35; Length 259;  
 Best Local Similarity 88.8%; Pred. No. 4.2e-41;  
 Matches 231; Conservative 0; Mismatches 28; Indels 1; Gaps 1;





Email: arkerlav@tigr.org  
 For clone availability, additional sequence and expression  
 information related to this EST, please check the TIGR Human Gene  
 Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
 Seq primer: M3 Reverse.

FEATURES  
 source  
 Location/Qualifiers  
 1. 216

/organism="Homo sapiens"  
 /db\_xref="ARCC (inhost):138014"  
 /db\_xref="taxon:9606"  
 /clone\_lib="Endometrial tumor"  
 /sex="female"  
 /dev\_stage="adult"  
 /note="Organ: endometrium; Vector: pBluescript SK-";  
 Site\_1: EcoRI; Site\_2: XhoI

BASE COUNT 45 a 65 c 63 g 38 t 5 others

ORIGIN

Query Match 13.9%; Score 176; DB 32; Length 216;  
 Best Local Similarity 97.2%; Pred. No. 4.3e-33;  
 Matches 176; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 2 GCAGCCGCGACTGCGAGCCCTGCGAGCGCGACTGTCATGGAAGCAATTGCTGCT 61  
 DB 36 GCAGCCGCGACTGCGAGCCCTGCGAGCGCGACTGTCATGGAAGCAATTGCTGCT 95  
 OY 62 CGGGCTCTGTCGTCGCGAGTGGTGTCTGACGCCGACACTGTTCCCAACTCTCT 121  
 DB 96 CGGGCTCTGTCGTCGCGAGTGGTGTCTGACGCCGACACTGTTCCCAACTCTCT 155  
 OY 122 ACACCATCGGGCTGGCTGCGAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 181  
 DB 156 ACACCATCGGGCTGGCTGCGAGTCTTGAGGCCGACCAAGAGCCAGGAGCCAGATGG 215  
 OY 182 T 182  
 DB 216 T 216

RESULT 12  
 LOCUS A1675815 188 bp mRNA EST 19-MAY-1999  
 DEFINITION wb97b12.x1 NCI\_CGAP\_Pri28 Homo sapiens cDNA clone IMAGE:2313599 3',  
 mRNA sequence.  
 ACCESSION A1675815  
 NID 94876295  
 VERSION A1675815.1 GI:4876295  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 TITLE Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Dec 20, 1995 this sequence version replaced gi:1133540.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: M. Bento Soares, Ph.D.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/bbrp/image/image.html](http://www.bio.llnl.gov/bbrp/image/image.html)  
 Seq primer: -40UP from G1bco.

FEATURES  
 source  
 Location/Qualifiers  
 1. 188

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_image="IMAGE:2313599"  
 /clone\_lib="NCI\_CGAP\_Pri28"  
 /sex="male"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"

/note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)  
 with a modified polylinker; Plasmid DNA from the  
 normalized library NCI\_CGAP\_Pri28 was prepared, and ss  
 circles were made in vitro. Following HAP purification,  
 this DNA was used as tracer in a subtractive hybridization  
 reaction. The driver was PCR-amplified cDNAs from a pool  
 of 5,000 clones made from the same library (clones  
 985608-986759, 1101192-1101959, and 1217928-1220615).  
 Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 43 a 31 c 61 g 53 t

ORIGIN

Query Match 13.2%; Score 167.2; DB 50; Length 188;  
 Best Local Similarity 92.9%; Pred. No. 5.6e-31;  
 Matches 169; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY 1084 ACCGAGGTNNAAGTCCGAGCCCTCTTCCTCAGACCCAGNGGTCCAAATGCCACTAG 1143  
 DB 188 ACCGAGGTNNAAGTCCGAGCCCTCTTCCTCAGACCCAGGTCCAAATGCCACTAG 129  
 OY 1144 ATTTTCCCTGNAACAGTCCCTCTTGAGNAGTTGACCCAACTTACAGTTGGTTT 1203  
 DB 128 ACTGCCCTGTACAGAGTCCCTCTTGAGNAGTTGACCCAACTTACAGTTGGTTT 69  
 OY 1204 TCATTTTNGTCCCTTCCCTAGATCCAGAAATTAAGTTTAAGANGNCAAAAAA 1263  
 DB 68 TCATTTTNGTCCCTTCCCTAGATCCAGAAATTAAGTTTAAGANGNCAAAAAA 9  
 OY 1264 AA 1265  
 DB 8 AA 7

RESULT 13  
 LOCUS AA595489 191 bp mRNA EST 18-SEP-1997  
 DEFINITION no34h04.s1 NCI\_CGAP\_Pri23 Homo sapiens cDNA clone IMAGE:1102615 3',  
 mRNA sequence.  
 ACCESSION AA595489  
 NID 92410839  
 VERSION AA595489.1 GI:2410839  
 KEYWORDS EST.  
 SOURCE human.

REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 TITLE Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 JOURNAL NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
 COMMENT National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 On Apr 14, 1993 this sequence version replaced gi:692625.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 cDNA Library Preparation: Stratagene, Inc.  
 cDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:







GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 23:25:10 : Search time 2910.9 Seconds  
(without alignments)  
1382.078 Million cell updates/sec

Title: US-09-030-606-173

Sequence: 1 GGCAAGCCCGCACTCGCAGCC.....AGAGANGNGCAAAAAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 679419 segs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_bai: \*  
2: gb\_baz: \*  
3: gb\_om: \*  
4: gb\_ov: \*  
5: gb\_pac: \*  
6: gb\_ph: \*  
7: gb\_pl: \*  
8: gb\_pl2: \*  
9: gb\_pl1: \*  
10: gb\_pr2: \*  
11: gb\_pr3: \*  
12: gb\_ro: \*  
13: gb\_st: \*  
14: gb\_sts: \*  
15: gb\_sy: \*  
16: gb\_un: \*  
17: gb\_vl: \*  
18: em\_fun: \*  
19: em\_hlg: \*  
20: em\_hum1: \*  
21: em\_hum2: \*  
22: em\_in: \*  
23: em\_om: \*  
24: em\_or: \*  
25: em\_ov: \*  
26: em\_pac: \*  
27: em\_ph: \*  
28: em\_pl: \*  
29: em\_ro: \*  
30: em\_sts: \*  
31: em\_sy: \*  
32: em\_un: \*  
33: em\_vl: \*  
34: gb\_hcg1: \*  
35: gb\_hcg2: \*  
36: gb\_in1: \*  
37: gb\_in2: \*  
38: em\_bai: \*  
39: em\_baz: \*  
40: em\_hum3: \*  
41: em\_hum4: \*  
42: gb\_pr4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	972.4	76.9	1347	42	AF113140	Homo sapi

2	553	43.7	5900	42	AF113141	AF113141 Homo sapi
3	474.8	37.5	4740	42	AF135023	AF135023 Homo sapi
4	474.8	37.5	4385	42	AF148532	AF148532 Homo sapi
5	452	35.7	1140	3	SSU76256	U76256 Sus scrofa
6	310	24.5	1237	12	AF019979	AF019979 Mus muscu
7	205.4	16.2	35197	11	AC005782	AC005782 Homo sapi
8	190.8	15.1	6592	9	HSCPT1175	X90780 Homo sapien
9	180.2	15.0	37785	9	HUMXRCC1G	L34079 Human XRCCL
10	194.6	14.6	8174	5	134189	134189 Sequence 5
11	174.2	13.8	281	10	HUMRSSA19	R03500 Human 37 bp
12	169	13.4	1665	9	D87943	D87943 Human DNA f
13	167.4	13.2	205251	11	AC005392	AC005392 Homo sapi
14	165.4	13.1	6592	10	HSCPT1175	X90780 Homo sapien
15	163.8	12.9	4033	9	AB006136	AB006136 Homo sapi
16	163.8	12.9	42710	42	AC006942	AC006942 Homo sapi
17	162.2	12.8	16689	10	HSAP011712	AF011712 Homo sapi
18	160.2	12.7	16689	10	HSAP011712	AF011712 Homo sapi
19	158.8	12.6	35465	11	AC005525	AC005525 Homo sapi
20	145.8	11.5	357	9	HUMSATVIS	M54997 Human DNA w
21	145.2	11.5	4060	5	I30049	I30049 Sequence 1
22	145.2	11.5	4067	10	SS1329	SS1329 AAVSI-adenov
23	144.4	11.4	37314	9	HUMAMDA	M63796 Human DNA f
24	140.6	11.1	4776	11	HSBCU352	U05681 Human proto
25	134.4	10.6	11084	11	AC006486	AC006486 Homo sapi
26	133.2	10.5	2403	10	HSNBNBN7	U31342 Human nucle
27	129	10.2	3840	10	HSAPCC2G	X05151 Human apoc-
28	125.8	9.9	4340	10	HUMAPOC2	M10612 Human apol1
29	113.2	8.9	4057	10	HUMAPOC2	U02658 Human apol1
30	112	8.9	4776	11	HSBCU352	U05681 Human proto
31	111.2	8.8	2711	9	AB012761	AB012761 Homo sapi
32	110.2	8.7	6500	42	AF095743	AF095743 Homo sapi
33	109.8	8.7	1438	5	E13202	E13202 Human gene
34	109.8	8.7	1419	9	D78203	D78203 Homo sapien
35	109.8	8.7	1506	10	HSUG2801	U52801 Human prote
36	109.8	8.7	1451	11	AF013988	AF013988 Homo sapi
37	106.2	8.4	732	5	I95869	I95869 Sequence 2
38	102	8.1	5574	11	AF055481	AF055481 Homo sapi
39	97.4	7.7	586	14	G41906	G41906 SHGC-56840
40	95	7.5	986	5	A42048	A42048 Sequence 1
41	95	7.5	969	11	HUMSERPROT	L33404 Human strat
42	90	7.1	881	12	RNTRYPB	X59013 Rat mRNA fo
43	89.6	7.1	729	11	HSHTS7	U92322 Homo sapien
44	89.2	7.1	893	4	PTTRYP	X56744 P.platessa
45	88.8	7.0	877	4	PMTRYPSIN	X82223 P.mage1lani

## ALIGNMENTS

RESULT 1

AF113140 LOCUS AF113140 1347 bp mRNA PRI 25-MAR-1999

DEFINITION Homo sapiens serine protease mRNA, complete cds.

ACCESSION AF113140

NTD 94512029

VERSION AF113140.1 GI:4512029

KEYWORDS

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1347)

AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.

TITLE Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)

MEDLINE 99179024

REFERENCE 2 (bases 1 to 1347)

AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and Wang,K.

TITLE Direct Submission

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,  
Bochell, WA 98021, USA  
FEATURES Location/Qualifiers  
source  
1. 1347  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13"  
1. 765  
/note="androgen regulated"  
/codon\_start=1  
/product="serine protease prostate"  
/protein\_id="AAD21580.1"  
/db\_xref="PID:94512030"  
/db\_xref="GI:4512030"  
translation= MATHGNPWFVFLYLILGVAGSLVSGSCQIINGEDCSPHSQPW  
QALVMELEFCGLVHPOWLSAHCFOHSYIIGLHSLLEADQEGSOMVBSLS  
VRHEYNRPLNDMLKIDESVSDTISISASCPAGNSCLVSGWLANGR  
MPTVLCVNVSVSEEVCSKLDPLVHSMCAGGHDOKDSCNDSGSLICNGYIQ  
GLVSFGAPCGGVGPVYTNLCFTMEIKETVQAS"

## CDS

1. 765  
/note="androgen regulated"  
/codon\_start=1  
/product="serine protease prostate"  
/protein\_id="AAD21580.1"  
/db\_xref="PID:94512030"  
/db\_xref="GI:4512030"  
translation= MATHGNPWFVFLYLILGVAGSLVSGSCQIINGEDCSPHSQPW  
QALVMELEFCGLVHPOWLSAHCFOHSYIIGLHSLLEADQEGSOMVBSLS  
VRHEYNRPLNDMLKIDESVSDTISISASCPAGNSCLVSGWLANGR  
MPTVLCVNVSVSEEVCSKLDPLVHSMCAGGHDOKDSCNDSGSLICNGYIQ  
GLVSFGAPCGGVGPVYTNLCFTMEIKETVQAS"

POLYA\_site 1327..1332  
BASE COUNT 269 a 489 c 334 g 255 t  
ORIGIN

Query Match 76.9% Score 972.4; DB 42; Length 1347;  
Best Local Similarity 87.0%; Pred. No. 3.5e-214;  
Matches 1149; Conservative 0; Mismatches 22; Indels 150; Gaps 2;

Oy 2 GCAGCCGCACTCGCAGCCCTGGCAGCGGGCAGCTGTCATGGAACGAATTGTCGT 61  
Db 110 GCAGCCGCACTCGCAGCCCTGGCAGCGGGCAGCTGTCATGGAACGAATTGTCGT 169  
Oy 62 CGGGGCTCTGTCATCCGCGAGTGGTGTCTCAGCCGCAACACTGTTCCAGAACTCT 121  
Db 170 CGGGGCTCTGTCATCCGCGAGTGGTGTCTCAGCCGCAACACTGTTCCAGAACTCT 229  
Oy 122 ACACCATGGGCTGGGCGCTGCACAGTCTTGAGCGGACCAAGCGAGGAGCGAGATGG 181  
Db 230 ACACCATGGGCTGGGCGCTGCACAGTCTTGAGCGGACCAAGCGAGGAGCGAGATGG 289  
Oy 182 TGGAGGCGAGCTCTCCGTACGCGACCAAGAGTACACAGACCTTGTCCGTTAAGCAG 241  
Db 290 TGGAGGCGAGCTCTCCGTACGCGACCAAGAGTACACAGACCTTGTCCGTTAAGCAG 349  
Oy 242 TCATCTCATCAAGTTGGAGCAATCCGTGTCCGAGTGTACACCATCGGAGCATCAGCA 301  
Db 350 TCATCTCATCAAGTTGGAGCAATCCGTGTCCGAGTGTACACCATCGGAGCATCAGCA 409  
Oy 302 TTGCTTCGAGTGGCCCTACCGCGGGGAACCTTGGCTGTTCTGGCTGGGCTGTGG 361  
Db 410 TTGCTTCGAGTGGCCCTACCGCGGGGAACCTTGGCTGTTCTGGCTGGGCTGTGG 469  
Oy 362 CGAGCGGTGAGCTCAAGGGGTGTGTCTGCCCTCTTCAAGAGAGTCTCTGTCCAGTGC 421  
Db 470 CGAAC----- 475  
Oy 422 GGGGGCTAGCCAGAGCTCTGGCTCCAGGCAAGATGCTTACCGTGTGCAATGGCTGAA 481  
Db 475 -----GGCAGAAATGCTTACCGGTGTGCAATGGCTGAA 506  
Oy 482 CGTTCGAGTGGTGTGAGAGAGTGTGACAGTAAAGCTTATGACCCGCTGTACACCCAG 541  
Db 507 CGTTCGAGTGGTGTGAGAGAGTGTGACAGTAAAGCTTATGACCCGCTGTACACCCAG 566  
Oy 542 CATGTTTCGCGCGCGGAGGCAAGACAGAGACTCTCTCAACGGTACTGTGGGG 601  
Db 567 CATGTTTCGCGCGCGGAGGCAAGACAGAGACTCTCTCAACGGTACTGTGGGG 626  
Oy 602 GCCCTGATCTCAACGGGTACTTGCAGGGCTTGTGTCTTTCGAAAAAGCCCGTGTGG 661  
Db 627 GCCCTGATCTCAACGGGTACTTGCAGGGCTTGTGTCTTTCGAAAAAGCCCGTGTGG 686

Oy 662 CCAAGTTGGCTGGCAGGTCGTCTACACCAACCTCTGCAAAATTCAGTGGATGAGAA 721  
Db 687 CCAAGTTGGCTGGCAGGTCGTCTACACCAACCTCTGCAAAATTCAGTGGATGAGAA 746  
Oy 722 AACGCTCCAGGCGAGTTAACTCTGGGAGCTGGGAACCCATGAAATTAAGCCCAATAACA 781  
Db 747 AACGCTCCAGGCGAGTTAACTCTGGGAGCTGGGAACCCATGAAATTAAGCCCAATAACA 806  
Oy 782 TCCTCGGAAGGAATTCAGGAATATCTGTTCCAGCCCTCTCTCCAGCCAGAGAGT 841  
Db 807 TCCTCGGAAGGAATTCAGGAATATCTGTTCCAGCCCTCTCTCCAGCCAGAGAGT 866  
Oy 842 CCAGGCCCCCGAGCCCTCTCTCCAGCAACCAAGGATCAGATCC----- 886  
Db 867 CCAGGCCCCCGAGCCCTCTCTCTCCAGCAACCAAGGATCAGATCC----- 926  
Oy 886 -----CCAGCCCT 894  
Db 927 CAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCCAGACCCAGAGAGTCCAGCCCT 986  
Oy 895 CCTCCCTCAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCTCCAGACCCAGAGTCC 954  
Db 987 CCTCCCTCAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCTCCAGACCCAGAGTCC 1046  
Oy 955 AGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCTCCAGACCC 1014  
Db 1047 AGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCTCCAGACCC 1106  
Oy 1015 GGGGTTGAGGCCCCCAACCCCTCTCTCTCAGAGTCCAGAGTCCAGACCCCAACCCCTC 1074  
Db 1107 GGGGTTGAGGCCCCCAACCCCTCTCTCTCAGAGTCCAGAGTCCAGACCCCAACCCCTC 1166  
Oy 1075 GTTCCCAAGACCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCC 1134  
Db 1167 GTTCCCAAGACCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCC 1226  
Oy 1135 GCCACCTGATTTTCCCTGACAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAG 1194  
Db 1227 GCCACCTGATTTTCCCTGACAGAGTCCAGAGTCCAGAGTCCAGAGTCCAGAGTCC 1286  
Oy 1195 GTTGGTTTTCATTTTNGTCCCTTCTCCCTAGATCCAGAAATTAAGTTAAGANG 1254  
Db 1287 GTTGGTTTTCATTTTNGTCCCTTCTCCCTAGATCCAGAAATTAAGTTAAGANG 1346  
Oy 1255 C 1255  
Db 1347 C 1347  
  
RESULT 2  
AF113141  
LOCUS AF113141 5900 bp DNA PRI 25-MAR-1999  
DEFINITION Homo sapiens serine protease gene, complete cds.  
ACCESSION AF113141  
NID 94512031  
VERSION AF113141.1 GI:4512031  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 5900)  
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and  
Wang,K.  
TITLE Molecular cloning and characterization of prostate, an  
androgen-regulated serine protease with prostate-restricted  
expression  
PROC. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)  
JOURNAL 99179024  
MEDLINE 2 (bases 1 to 5900)  
REFERENCE  
Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gelinas,R., Hood,L. and  
Wang,K.  
TITLE Direct Submission



Db 2841 CCACAGCTCTCTACACCATGGGGCTGGGCGCCGACAGCTTGTAGCGCCAGCAAGAGCCAGG 2900  
QY 171 GAGCCAGATGTGGAGGCCAGCCTCTCCGTACGGCACCAGAGTAAACAGACCTTGTCT 230  
Db 2901 GAGCCAGATGTGGAGGCCAGCCTCTCCGTACGGCACCAGAGTAAACAGACCTTGTCT 2960  
QY 231 CGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGCTGACACATCCG 290  
Db 2961 CGGTACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGCTGACACATCCG 3020  
QY 291 GAGCATCAGCATTCCTCCAGAGTCCCTTACCCGCGGGAATCTTGCCTGTTTGGCTG 350  
Db 3021 GAGCATCAGCATTCCTCCAGAGTCCCTTACCCGCGGGAATCTTGCCTGTTTGGCTG 3080  
QY 351 GGGTCTGCTGGGCAAGGCTGAGCTCAGGGGTGTGTCTGCGCCCTTCAAGAGGCTCTC 410  
Db 3081 GGGTCTGCTGGGCAAGGCTGAGCTCAGGGGTGTGTCTGCGCCCTTCAAGAGGCTCTC 3140  
QY 411 TGCCAGTGGCGGGGCTGACCCAGAGCTCTGCGTCCAGAGCAATGCTACCGTGTG 470  
Db 3141 TGCCAGTGGCGGGGCTGACCCAGAGCTCTGCGTCCAGAGCAATGCTACCGTGTG 3200  
QY 471 CAGTGGCTGACGTCTGCTGCTGTCTGAGAGAGTCTGCAATGCTATGACCGCTG 530  
Db 3201 CAGTGGCTGACGTCTGCTGCTGTCTGAGAGAGTCTGCAATGCTATGACCGCTG 3260  
QY 531 TACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCAAGACCAAGAGACTCTGCAAG 588  
Db 3261 TACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCAAGACCAAGAGACTCTGCAAG 3318

RESULT 4  
AF148532  
LOCUS AF148532 4385 bp DNA PRI 09-JUN-1999  
DEFINITION Homo sapiens kallikrein 4 (KLK4) gene, complete cds.  
ACCESSION AF148532  
NID 95020095  
VERSION AF148532.1 GI:5020095  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 4385)  
AUTHORS Stephenson, S.A., Verity, K., Ashworth, L. and Clements, J.A.  
TITLE Localization of a new prostate specific antigen-related serine  
protease gene, KLK4, is evidence for an expanded human kallikrein  
(KLK) gene family cluster on chromosome 19q13.3-13.4  
JOURNAL Unpublished  
AUTHORS Stephenson, S.A., Ashworth, L. and Clements, J.A.  
TITLE Direct Submission  
JOURNAL Submitted (03-May-1999) School of Life Science, Queensland  
University of Technology, George Street, Brisbane, Queensland 4001,  
Australia

FEATURES  
source 1. 4385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13.3-q13.4"  
join(<551..611,1875..2037,2459..2709,2793..2929,  
4200..>4352)  
/gene="KLK4"  
/product="kallikrein 4"  
/gene="KLK4"  
join(551..611,1875..2037,2459..2709,2793..2929,4352)  
CDS  
gene  
CDS  
/gene="KLK4"  
/function="serine protease"  
/codon\_start=1  
/product="kallikrein 4"

/protein\_id="AAD38019.1"  
/db\_xref="PId:95020096"  
/db\_xref="GI:5020095"  
/translation="MATAGNPMGFFLVLILGNVLSVSGSCSITINGEDCSFHSQPM  
QALVMENELFCSSVLVHPQVLSAALCFQSTYTGILSLSDQEPGSDQWEASLS  
VRHPEYRNPFLANDLMLIKLDESVESDITRSISIAQCPYAGNSCLVSGWGLANR  
MPVLQGVNVSVEEYCSKLYDPLVHPSFECAGGGODKDCNSGSGPLICNGYLO  
GLVFEGRAPCGQVGPVYTNLCKFTETMIERTVQAS"  
BASE COUNT 1013 a 1184 c 1192 g 996 t  
ORIGIN

Query Match 37.5%; Score 474.8; DB 42; Length 4385;  
Best Local Similarity 99.6%; Pred. No. 9.4e-100;  
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 111 CCAGAACTCTACACCATGGGCTGGGCTGCACAGTCTTGAGGCCAGCAAGAGCCAGG 170  
Db 2453 CCACAGCTCTACACCATGGGCTGGGCTGCACAGTCTTGAGGCCAGCAAGAGCCAGG 2512  
QY 171 GAGCAGATGGTGGAGGCGAGCTCTCCGTACGGCAGCAGATCAACAGACCTTGTCT 230  
Db 2513 GAGCAGATGGTGGAGGCGAGCTCTCCGTACGGCAGCAGATCAACAGACCTTGTCT 2572  
QY 231 CGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGCTGACACCATCCG 290  
Db 2573 CGCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTGTCGAGCTGACACCATCCG 2632  
QY 291 GAGCATCAGCATTCCTCCAGAGTCCCTTACCCGCGGGAATCTTGCCTGTTTGGCTG 350  
Db 2633 GAGCATCAGCATTCCTCCAGAGTCCCTTACCCGCGGGAATCTTGCCTGTTTGGCTG 2692  
QY 351 GGGTCTGCTGGGCAAGGCTGAGCTCAGGGGTGTGTCTGCGCCCTTCAAGAGGCTCTC 410  
Db 2693 GGGTCTGCTGGGCAAGGCTGAGCTCAGGGGTGTGTCTGCGCCCTTCAAGAGGCTCTC 2752  
QY 411 TGCCAGTGGCGGGGCTGACCCAGAGCTCTGCGTCCAGAGCAATGCTACCGTGTG 470  
Db 2753 TGCCAGTGGCGGGGCTGACCCAGAGCTCTGCGTCCAGAGCAATGCTACCGTGTG 2812  
QY 471 CAGTGGCTGACGTCTGCTGCTGTCTGAGAGAGTCTGCAATGCTATGACCGCTG 530  
Db 2813 CAGTGGCTGACGTCTGCTGCTGTCTGAGAGAGTCTGCAATGCTATGACCGCTG 2872  
QY 531 TACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCAAGACCAAGAGACTCTGCAAG 588  
Db 2873 TACCAACCCAGCATGTTCTGCGCGCGGCGGAGGCAAGACCAAGAGACTCTGCAAG 2930

RESULT 5  
SSU76256  
LOCUS SSU76256 1140 bp mRNA MAN 03-JAN-1998  
DEFINITION Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,  
complete cds.  
ACCESSION SSU76256  
NID 92737920  
VERSION U76256.1 GI:2737920  
KEYWORDS  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
REFERENCE 1 (bases 1 to 1140)  
AUTHORS Simmer, J.P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J.,  
Margolis, H.C., Shimizu, M., Hu, C.-C. and Bartlett, J.D.  
TITLE Purification, Characterization and Cloning of Enamel Matrix Serine  
Proteinase 1  
JOURNAL Unpublished  
AUTHORS Simmer, J.P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J.,  
Margolis, H.C., Shimizu, M., Hu, C.-C. and Bartlett, J.D.  
TITLE Direct Submission  
JOURNAL Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas

Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78284-7888, USA

## FEATURES

Location/Qualifiers

## source

1..1140 /organism="Sus scrofa"

/db\_xref="taxon:9823"

## sig\_peptide

69..140

/note="the signal peptide is cleaved after Ala24 and the preprotein is secreted into the developing enamel matrix"

## CDS

69..833

/note="EMSP1"

/codon\_start=1

/product="enamel matrix serine proteinase 1 precursor"

/protein\_id="AAB94638.1"

/db\_xref="PIR:g2737921"

/db\_xref="GI:2737921"

/translation="MTAGNPMGFLGLIGTSLAFINGSHIINGEDCNPHSPWQALFLDEDFCGGLVHPOMVLSAHCPSNTYIGLHNLPEQPGSQMDEASISIOHREYEPNPMANDLMLIKLESYLSIDTNTSVSGDPTPGDSCISVGMRLASRLPVLOCNVISVAEEYKARVPYHPSMFCAGDQDSDHSDGDFLNCISLQGLVFGAOCGQPNVPEVYTNLCKFTDITTIQAS"

159..830

## mat\_peptide

159..830

/note="the preprotein is cleaved following His30 activating the serine proteinase; the active protein has an apparent molecular weight of 34 kDa and a derived molecular weight of 24 kDa"

/product="enamel matrix serine proteinase 1"

/order(177..179,549..551)

/note="disulfide bond"

/order(234..236,282..284)

/note="disulfide bond"

/join(279..281,414..416,687..689)

/note="encodes catalytic triad"

405..407

/note="encodes potential glycosylation location; yields a blank cycle during protein sequencing"

/order(408..410,789..791)

/note="disulfide bond"

/order(510..512,705..707)

/note="disulfide bond"

/order(675..677,750..752)

/note="disulfide bond"

BASE COUNT 245 a 366 c 295 g 234 t

## ORIGIN

ORIGIN

## Query Match

Best Local Similarity 74.1%; Score 452; DB 3; Length 1140;

Matches 640; Conservative 0; Mismatches 140; Indels 84; Gaps 2;

Query 2 GCAGCCGCACTGCGAGCCCTGGGAGGCGGCGATGTCATGGAAGAATGTTGCTCT 61

Db 178 GCAACCCCTCCAGCCCTGGGAGGCGGCGATGTTCTCTGGAAGACGACTTTTCTCG 237

Query 62 CGGGCGCTCGGTGATCCGAGTGGGTGCTGACGCGGCACTGTTCCAGAACTCT 121

Db 238 GGGGCGTCTGTCATCCGAAATGGGTGCTGACGCGGCACTGTTCCAAATTTCT 297

Query 122 ACACCATGCGGCTGGGCTGACAGTCTTGGAGCGGACCAAGACGAGGATGG 181

Db 298 ACACCATGCGGCTGGGCTGACATCTTGAAGCCGGAAGAACCGGAGCCAGATGA 357

Query 182 TGAAGCCAGGCTCTCGGTACGCGACCGAGTACAGACACCTTGTCTGCTAAGAC 241

Db 358 TGAAGCCAGGCTCTCGGTACGCGACCGAGTACAGACACCTTGTGCTAAGAC 417

Query 242 TCATGCTCATAGTGGAGCAATCCGTGCTGAGTCTGACACCATCCGAGCATCACA 301

Db 418 TCATGCTCATAGTGGAGCAATCCGTGCTGAGTCTGACACCATCCGAGCATCACA 477

Query 302 TTGCTTGCAGTGGCCCTGACGCGGAGCACTTGGCTGTTGCTGGGCTGTGG 361

Db 478 TCGTCTCCCATGCGCGGACCCCTGGGGAATCTTGGCTGTTCCGGCTGGGCTGG 537

Query 362 CGAAGGTGAGCTACGCGGTGTGTCTGCCCTTTCAAGAGTCTCTGCCCCAGTCCG 421

Db 538 C----- 539

Query 422 GGGGGTGAACCAAGCTCTGCGTCCGAGCAGAAATGCTACGCTGCGAGTGGTGA 481

Db 539 -----CAATGGCAGACTGCCCAAGTGTGAGTGTGGA 574

Query 482 CGTGTGCTGTGTGTGAGAGTGTGCTGACAGTATGATGACCCGCTGACACCCAG 541

Db 575 CATCTCGGTGCTGTGAGAGTGTGCTGACAGGCGGCTGACGCGCCGCTGACACCCAG 634

Query 542 CATTTTGTGCGCGCGGAGGAGGCAAGACAGAAAGATCTCTGACAGCTGACTGTGGGG 601

Db 635 CATTTTGTGCGCGCGGAGGAGGCAAGACAGAAAGATCTCTGACAGCTGACTGTGGGG 694

Query 602 GCCCTGATCTGCAACGGGACTGTCAGAGGCGCTGTCTTTGCGAAAGACCCGCTGG 661

Db 695 CCCCTCATCTGCAATGGTCCCTGCAAGGCGCTGTCTCTTGGACAGCCAGTGTGG 754

Query 662 CCAAGTTGGGTGTCAGAGTGTGACACCAACTCTGCAAAATTCAGTGTGATAGAA 721

Db 755 CCAACCCAGAGTGGCAGGCGCTGACACCAACTCTGCAAAATTCAGTGTGATAGAA 814

Query 722 AACGTCAGGCGGATTAATCTGTGGGACTGGGAACCATGAATTAACCCCAATACA 781

Db 815 GACATCATGAGCCGATTAATCTGTGGGACTGGGAACCATGAATTAACCCCAATACA 874

Query 782 TCGTGGGAAGAAATTCAGAAATATGTTCCAG-CCCTGCTCCCTGAGGCGGAG 840

Db 875 CTCGTCTGAAGAAATTCAGAAATATGTTCCAG-CCCTGCTCCCTGAGGCGGAG 934

Query 841 TCCAGGCCCGGAGCCCTCTCC 864

Db 935 TCTAGACACCCCGGAGCCCTCC 958

## RESULT

6

## AF019979

## LOCUS

AF019979

## DEFINITION

Mus musculus enamel matrix serine proteinase 1 precursor, mRNA,

## ACCESSION

AF019979

## NID

94090846

## VERSION

AF019979.1

## KEYWORDS

house mouse,

## SOURCE

Mus musculus

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

## REFERENCE

1 (bases 1 to 1237)

## AUTHORS

Simmer,J.

## TITLE

Enamel Matrix Serine Proteinase 1 (EMSP1)

## JOURNAL

2 (bases 1 to 1237)

## REFERENCE

Direct Submission

## TITLE

Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas

## JOURNAL

Health Science Center at San Antonio, 7703 Floyd Curl Drive, San

Antonio, TX 78284-7888, USA

## FEATURES

Location/Qualifiers

1..1237

/organism="Mus musculus"

/strain="Swiss-Webster"

/db\_xref="taxon:10090"

/cell\_type="ameloblast"

52..126

/function="degradation of the enamel matrix during enamel

maturation"

/note="EMSP1"

/codon\_start=1





```

misc-feature
/tpc_family="GC-rich"
338..430
/Note="BLASTX similarity to 008509 (343..373): match:
0.51, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir|S39983
eps8 protein - mouse >g1309217 (L21671) Eps8 [Mus]"
338..412
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
517..660
misc-feature
/Note="BLASTX similarity to 008509 (368..415): match:
0.52, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir|S39983
eps8 protein - mouse >g1309217 (L21671) Eps8 [Mus]"
517..665
misc-feature
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000--DDS
similarity to overlapping ESTs:~(611..665) R34158
yh83h09.s1 Homo sapiens cDNA clone 136385 3' similar to
SP:S39983 S39983 EPS8 PROTEIN:(1..55): 100%
identity:~(621..665) R80263 y196h06.s1 Homo sapiens cDNA
clone 147131 3' (1..45): 100% identity."
complement(1145..1258)
/tpc_family="FLAM_C"
1441..1582
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 75.000--DDS
similarity to overlapping ESTs:~R34158 yh83h09.s1 Homo
sapiens cDNA clone 136385 3' similar to SP:S39983 S39983
EPS8 PROTEIN:(56..194): 94% identity:~R80263 y196h06.s1
Homo sapiens cDNA clone 147131 3' (46..186): 94%
identity:~H13945 y108d11.s1 Homo sapiens cDNA clone
148149 3' similar to SP:S39983 S39983 EPS8 PROTEIN:
(21..161): 94% identity."
complement(1621..1652)
/tpc_family="GC-rich"
1679..1707
/Note="DSS similarity to overlapping ESTs:~R34158
yh83h09.s1 Homo sapiens cDNA clone 136385 3' similar to
SP:S39983 S39983 EPS8 PROTEIN:(195..223): 97%
identity:~R80263 y196h06.s1 Homo sapiens cDNA clone
147131 3' (187..215): 97% identity:~H13945 y108d11.s1
Homo sapiens cDNA clone 148149 3' similar to SP:S39983
S39983 EPS8 PROTEIN:(162..190): 93% identity."
complement(1880..1975)
/tpc_family="MSR1"
3160..3240
/tpc_family="MIR"
complement(3241..3527)
/tpc_family="AluSc"
3607..3702
/tpc_family="MIR"
3904..4036
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 99.000--DDS
similarity to overlapping ESTs:~(3904..3988) R80263
y196h06.s1 Homo sapiens cDNA clone 147131 3' (216..300):
96% identity:~(3904..4036) H13945 y108d11.s1 Homo sapiens
cDNA clone 148149 3' similar to SP:S39983 S39983 EPS8
PROTEIN:(191..323): 95% identity."
3989..4036
/Note="DSS similarity to R34158 yh83h09.s1 Homo sapiens
cDNA clone 136385 3' similar to SP:S39983 S39983 EPS8
PROTEIN:(224..271): 100% identity."
4124..4255
misc-feature
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 90.000--DDS
similarity to:~(4124..4215) R34158 yh83h09.s1 Homo
sapiens cDNA clone 136385 3' similar to SP:S39983 S39983
EPS8 PROTEIN:(272..364): 94% identity:~(4124..4166)
H13945 y108d11.s1 Homo sapiens cDNA clone 148149 3'
similar to SP:S39983 S39983 EPS8 PROTEIN:(324..336): 100%
identity."

misc-feature
4455..4581
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 2, quality: excellent, score: 95.000"
4777..4951
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 2, quality: excellent, score: 93.000"
complement(4817..4878)
/tpc_family="GC-rich"
5366..5498
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 2, quality: excellent, score: 93.000--DDS
similarity to:~(5367..5498) H13944 y108d11.r1 Homo sapiens
cDNA clone 148149 5' similar to SP:S39983 S39983 EPS8
PROTEIN:(368..236): 93% identity."
5591..5677
misc-feature
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 89.000--DDS
similarity to:~(5591..5825) H13944 y108d11.r1 Homo sapiens
cDNA clone 148149 5' similar to SP:S39983 S39983 EPS8
PROTEIN:(235..1): 99% identity."
complement(5619..5988)
/Note="BLASTN similarity to A1126400 (11..380): match:
0.99, score: 5.0e-144; database searched: month.na;
gc55f11.x1 Soares.Placenta.Bco9weeks_2bHP8to9W Homo
sapiens cDNA clone IMAGE:1713549 3', mRNA sequence [Homo
sapiens]"
complement(5691..5961)
/Note="DSS similarity to R34157 yh83h09.r1 Homo sapiens
cDNA clone 136385 5'. Score: 513 identity: 268/270
(99%)."
5852..5983
/Note="BLASTX similarity to D25793 (1..132): match: 0.98,
score: 2.8e-45; database searched: est; Human colon
3'directed kb01 cDNA, HUKGS04161, clone cm1029."
complement(6867..6897)
/tpc_family="GC-rich"
7429..8057
/tpc_family="LINE2"
complement(8297..8602)
/tpc_family="AluX"
8979..9294
/Note="DSS similarity to T23821 seq1081 Homo sapiens cDNA
clone b4HB3A-Cot51.5-HAP-Ft-23 3'. Score: 597 identity:
311/319 (97%)."
Additional EST matches: A1125316, T33476
and many others"
complement(9535..9621)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 92.000"
complement(9677..9763)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"
complement(10285..10389)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
complement(10580..10651)
/tpc_family="GC-rich"
complement(10685..10725)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 85.000"
complement(10810..10931)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score:
79.000--(10928..10906) DDS similarity to R35625 y966g06.r1
Homo sapiens cDNA clone 38310 5' (319..341): 78%
identity."
complement(11079..11139)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 1, quality: excellent, score: 83.000--DDS
similarity to R35625 y966g06.r1 Homo sapiens cDNA clone
38310 5' (254..318): 91% identity."
complement(11216..11296)
/Note="predicted exon, program: grall2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000--DDS

```



OY	783	CCTGGGAAGAAATATCATGATATCTGTCCAGCCCTCTCTCCCTCAGGCGCAGGAGTC	842	/note="exon 4; G00-120-737"
				26043..26117
				/partial
Db	2949	CCTCTCTCCTTGAAGACCTTGAAAGTCCAGGCCCTCTCTCTCTGAGACCTGGAGTC	3008	/gene="XRCC1"
				/note="exon 5; G00-120-737"
				26214..26325
				/partial
OY	843	CAGGCCCCAGCCCTCTCTCTCTCAACAGAGGTACAGATCCCAAGCCCTCTCTCTC	902	/gene="XRCC1"
				/note="exon 6; G00-120-737"
Db	3009	CAGG-CCCCAGCCCTCTCTCTCAGACCTTGAGTCCAGGCCCTCAAGCCCTCTCTCTC	3067	complement(26342..26484)
				/partial
OY	903	AGACCCAGAGTCCAGACCCCCAGCCCTCTCTCTCAGACCCAGAGT------CCA	955	/gene="XRCC1"
				/note="exon 6; G00-120-737"
Db	3068	AGACCTCTGAGTCCAG-CCCCCAGCCCTCTCTCTCAGACCTTGAGTCCAGGCCCA	3126	complement(26342..26484)
				/partial
OY	956	GGCCCTCTCTCAGACCCAGAGTCCAGACCCCCAGCCCTCTCTCTCAGACCCAG	1015	/note="human chromosome 19-specific tandem repeat, p6670"
				26535..26744
Db	3127	GGCCCTCTCTCTCAGACCCAGAGTCCAG-TCGCCAGCCCTCTCTCTCAGACCCAG	3184	/partial
				/gene="XRCC1"
OY	1016	GAGGTGAGGCCCCCAACCCCTCTCTCTCAGAGTCCAGAGTCCAGGCCCAACCCCTG	1075	/note="exon 7; G00-120-737"
				26818..26929
Db	3185	GATGAGAGCCCTCAGGCCCTCTCTCTCAGACCTTGAGTCC-AGGCCCCAGCCCTC	3243	/partial
				/gene="XRCC1"
OY	1076	TTCCCCAGACCCAGAGGTGAGTCCAGCCCTCTCTCTCAGACCCAGNGTCCATG	1135	/note="exon 8; G00-120-737"
				complement(27008..27410)
Db	3244	TCCCTCAGACCTTGAGTCCAGGCCCCAGCCCTCTCTCTCAGACCCAGAGTCCAGG	3303	/partial
				/gene="XRCC1"
OY	1136	CCACCTAGATTTTCCT 1152		/note="human chromosome 19-specific tandem repeat, p6670"
				27451..27709
Db	3304	CCACGCCCTCTCTCTCT 3320		/partial
				/gene="XRCC1"
				/note="exon 9; G00-120-737"
				28039..28155
				/partial
				/gene="XRCC1"
				/note="exon 10; G00-120-737"
				32609..32702
				/partial
				/gene="XRCC1"
				/note="exon 11; G00-120-737"
				32859..32991
				/partial
				/gene="XRCC1"
				/note="exon 12; G00-120-737"
				complement(33001..33360)
				/partial
				/gene="XRCC1"
				/note="human chromosome 19-specific tandem repeat, p6670"
				33480..33534
				/partial
				/gene="XRCC1"
				/note="exon 13; G00-120-737"
				33629..33768
				/partial
				/gene="XRCC1"
				/note="exon 14; G00-120-737"
				33533..33543
				/partial
				/gene="XRCC1"
				/note="exon 15; G00-120-737"
				33589..33595
				/partial
				/gene="XRCC1"
				/note="exon 16; G00-120-737"
				36082..36195
				/partial
				/gene="XRCC1"
				/note="exon 17; G00-120-737"
				complement(33001..33360)
				/partial
				/gene="XRCC1"
				/note="exon 18; G00-120-737"
				36195..36208
				/partial
				/gene="XRCC1"
				/note="exon 19; G00-120-737"
				36208..36214
				/partial
				/gene="XRCC1"
				/note="exon 20; G00-120-737"
				36214..36220
				/partial
				/gene="XRCC1"
				/note="exon 21; G00-120-737"
				36220..36226
				/partial
				/gene="XRCC1"
				/note="exon 22; G00-120-737"
				36226..36232
				/partial
				/gene="XRCC1"
				/note="exon 23; G00-120-737"
				36232..36238
				/partial
				/gene="XRCC1"
				/note="exon 24; G00-120-737"
				36238..36244

OY	807	CTGTTCCAGAGCCCTCTCTCCCTCAGAGCCAGAGTCCAGAGCCCTCCCTCTCCCTC	866
Db	27335	CAGGCCCCAGGCCCTCTCCCTCCTCAGACCAGAGGGTCCAGGCCCTCTCTCTCTC	27276
OY	867	AAACCAAGG-GTACAGATCCCGAGCCCTCTCTCCTCAGACCAGAGTCCAGAGCCCTCC	925
Db	27275	AGACCCAGAGAGTCCAGGGCCCCCAGGCCCTCTCTCCTCAGAGCCAGAGTCCAG-GGCCCC	27217
OY	926	AGCCCTCTCTCTCTCTAGAGCCAGAGTCCAG-----CCCTCTCTCCTCAGACCAG	977
Db	27216	AGCCCTCTCTCTCTCTCAGACCAGGGATCCAGGCCCCCAACCCCTCTCTCTCAGACCAG	27157
OY	978	GAGTCAGAGCCCCCAGGCCCTCTCTCTCTCAGACCAGAGGGTTCAGAGCCCCCAACCCCTC	1037
Db	27156	GGGTCCAG-GCTCCAGGCCCTCTCTCTCTCAGACCAGAGGGTTCAGAGCCCCCTCAGGCCCTC	27098
OY	1038	CTCTCTCAGAGTCAGAGTCCAGGCCCCCAACCCCTCTCTCTCTCCCGAGACCAGAGGTNNAG	1097
Db	27097	CTCCCTCGGATTCAGAGTAATCCAGAGGCCCCAGGCCCTCTCTCTCTCAGACCAGAGGGTCCAG	27038
OY	1098	G-TCCCAAGCCCTCTCTCCCTCAGACCAGAGGGTTCATTCACCTAGATTTCCTCTGNAC	1156
Db	27037	GCCCCACAGCCCTCTCTCTCTCAGATTTCAGAGCAGAGGCTCCAGGCCCGCTCTCCCTTAGG	26978
OY	1157	ACAGTGGCCCTCTGTGGNAGTTCAGACCACCACTTACCACTTGGTT 1201	
Db	26977	ACACAGGAGACAGGCCCTCTGTGGAAACAGAGGATCTAATTAGCT 26933	

RESULT	10				
LOCUS	134189/c				
DEFINITION	Sequence 5 from patent US 5595900.				
ACCESSION	134189	8174 bp	DNA	PAT	27-JAN-1997
NID	g1824980				
VERSION	134189.1	GI:1824980			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 8174)				
TITLE	Lowe, J. B.				
FEATURES	Methods and products for the synthesis of oligosaccharide structures on glycoproteins, glycolipids, or as free molecules, and for the isolation of cloned genetic sequences that determine these structures				
JOURNAL	Patent: US 5595900-A 5 21-JAN-1997;				
FEATURES	Location/Qualifiers				
SOURCE	1..8174				
BASE COUNT	1628 a 2229 c 2322 g 1995 t				
ORIGIN	/organism="unknown"				

Query Match	14.6%	Score 184.6	DB 5	Length 8174
Best Local Similarity	78.2%	Pred. No. 5.1e-33		
Matches 272; Conservative	0	Mismatches 64	Indels 12	Gaps 4

[illegible]

QY 1026 CCCCACCCCTCCCTCCCTTCAGAGTCCAGAGTCCAAAGCCCCCAACCCCTGGTCCCAAG 1085

Db 787 -CCAGACCTCTCTCCCTCCAGACCAGAGATCCAGCCCCCAACCCCTCTCTCCCTCAAGC 729

QY 1086 CCAAGAGTNNAGTCCCAAGCCCTCTTCCCTCAGACCCAGAGNGTCCAA 1133

Db 728 CCAAGAGTCCAGAGCCCAACCCCTCTCTCCCTCAGACCAAGAGGGCTTA 681

RESULT	11				
HMRSSA19					
LOCUS					
DEFINITION	HMRSSA19	281 bp	DNA	PII	09-MAY-1996
ACCESSION	Human 37	bp	minisatellite repeats,	specific to chromosome 19,	
NID	K03500				
VERSION	9337672				
KEYWORDS	K03500.1	GI:337672			
SOURCE	repeat:region.				
ORGANISM	Homo sapiens (clone library: COS 4C)				DNA.
	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
	Eutheria; Primates; Catarrhini; Homnidae; Homo.				
REFERENCE	1 (bases 1 to 281)				
AUTHORS	Das,H.K., Jackson,C.L., Miller,D.A., Leif,T. and Breslow,J.L.				
TITLE	The human alpha1proteobin C-II gene sequence contains a novel				
	chromosome 19-specific minisatellite in its third intron				
JOURNAL	J. Biol. Chem.	262 (10),	4787-4793		(1987)
MEDLINE	87165892				

FEATURES	source	Location/Qualifiers
		1..281
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/clone_lib="COS 4C"
		/cell_line="jy"
		1..37
satellite		/note="sat rpt copy A"
		38..75
satellite		/note="sat rpt copy B"
		76..113
satellite		/note="sat rpt copy C"
		114..151
satellite		/note="sat rpt copy D"
		152..188
satellite		/note="sat rpt copy E"
		189..225
satellite		/note="sat rpt copy F"
		226..262
satellite		/note="sat rpt copy G"
		263..281
satellite		/note="sat rpt partial copy H"
		58 a 147 c 44 g 32 t
BASE COUNT		
ORIGIN	Chromosome 19, long arm.	

Query Match	13.8%;	Score 174.2;	DB 10;	Length 281;
Best Local Similarity	82.4%;	Pred. No. 1.5e-30;		
Matches 224;	Conservative 0;	Mismatches 38;	Indels 10;	Gaps 2;

QY	876	GTAAGATATCCCAAGCCCCCTCTCTCCCTCAACACCCAGAGAGTGCAGACCCCCAGCCCTCT	935
Db	2	GTCAAGAGCCCCCAGCCCCCTCTCTCTCAACACCCAGAGAGTCAAGAACCCCCAGCCCTCTCT	61
QY	936	CCCTCAGACCCAGAGAGT-----CGAGCCCTCTCTCMTCAGACCCAGAGAGTCCGA	986
Db	62	CCCTCAGACCCAGAGAGTCAAGAACCCCCAGCCCTCTCTCTCTCAACCCAGAGAGTCAAGA	121
QY	987	CCCCCAGCCCTCTCTCTCTCAGACCCAGAGGGTTGAGGCCCCCAACCCCTCTCTTTAG	1046
Db	122	CCCCCAGCCCTCTCTCTCTCTCACTCATGAGTCCAGACCCCCAGCCCTCTCTCTCTAG	181
QY	1047	AGTGAAGGTTCAGAGCCCCCAACCCCTCTGTTCCCGACCCAGAGGTNNAG-GTCCAGC	11059
Db	182	ACCAGGAGGTTCAGAGCCCCCAGCCCTCTCTCTCTCTAGACCCAGAGGTTCAGAGCCCGC	241

OY 1106 CCCTCTTCNTCAGACCCAGAGGTCACATGCC 1137  
 Db 242 CCCTCTTCNTCAGACCCAGAGGTCACATGCC 273

RESULT 12  
 Db7943/c 1665 bp DNA PRI 29-JUL-1997  
 DEFINITION Human DNA for alpha(1,2)fucosyltransferase, intron 1.  
 ACCESSION D87943  
 MID 91842173  
 VERSION D87943.1 GI:1842173  
 KEYWORDS FUT1: alpha (1, 2) fucosyltransferase.  
 SOURCE Homo sapiens peripheral leukocytes DNA.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 1665)

REFERENCE 1  
 AUTHORS Kimura, H.  
 TITLE Direct Submission  
 JOURNAL Submitted (17-SEP-1996) to the DDBJ/EMBL/GenBank databases. Hiroshi Kimura, Kurume University, School of Medicine, Department of Legal Medicine, 67 Asahimachi, Kurume, Fukuoka 830, Japan (E-mail: hkimura@kurume.ktarn.or.jp, Tel: 0942-31-7554, Fax: 0942-31-7700)

REFERENCE 2 (sites)  
 AUTHORS Koda, Y., Soejima, M. and Kimura, H.  
 TITLE Structure and expression of H-type GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase gene (FUT1). Two transcription start sites and alternative splicing generate several forms of FUT1 mRNA  
 JOURNAL J. Biol. Chem. 272 (11), 7501-7505 (1997)  
 MEDLINE 97207318

REFERENCE 3 (sites)  
 AUTHORS Larsen, R.D., Ernst, L.K., Nair, R.P. and Lowe, J.B.  
 TITLE Molecular cloning, sequence, and expression of a human GDP-L-fucose:beta-D-galactoside 2-alpha-L-fucosyltransferase cDNA that can form the H blood group antigen  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 87 (17), 6674-6678 (1990)  
 MEDLINE 90370848

REFERENCE 4 (sites)  
 AUTHORS Kelly, R.J., Ernst, L.K., Larsen, R.D., Bryant, J.G., Robinson, J.S. and Lowe, J.B.  
 TITLE Molecular basis for H blood group deficiency in Bombay (Oh) and para-Bombay individuals  
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (13), 5843-5847 (1994)  
 MEDLINE 94286534

COMMENT Sequence updated (10-Feb-1997) by: Hiroshi Kimura.

FEATURES  
 source location/Qualifiers  
 1..1665  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="19q13.3"  
 /tissue\_type="peripheral leukocytes"  
 1..1665  
 /gene="FUT1"  
 1..1665  
 /gene="FUT1"  
 /number=1  
 BASE COUNT 349 a 382 c 513 g 421 t  
 ORIGIN

Query Match 13.48; Score 169; DB 9; Length 1665;  
 Best Local Similarity 76.98; Pred. No. 2; 1e-29;  
 Matches 283; Conservative 0; Mismatches 70; Indels 15; Gaps 6;

OY 795 ATTCAGGATATATGTCAGAGCCCTCTCTCCAGAGGAGCCAGAGCCAGC 854  
 Db 381 ACTTAGAGTCAGAGGTCCTCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 322  
 OY 855 CCCT 913  
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 321 CCCT 262  
 OY 914 TCCAGAGCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 965  
 Db 261 TCCAG-CTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 203

OY 966 CMTAGAGCCAGAGTCAGAGCCCGCCCTCTCTCTCTCTCTCTCTCTCTCT 1025  
 Db 202 CTTAGAGCCAGAGGAGTCAGAGTCCTCTCTCTCTCTCTCTCTCTCTCTCT 143

OY 1026 CCCCAGCCCT 1085  
 Db 143 -CCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 87

OY 1086 CCAGAGTNNAGG-TCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1144  
 Db 86 CCAGAGTCAGAGGAGCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 27

OY 1145 TTTTCCCT 1152  
 Db 26 TCTCTCTCT 19

RESULT 13  
 AC005392 205251 bp DNA PRI 04-DEC-1998  
 LOCUS AC005392/c  
 DEFINITION Homo sapiens chromosome 19, C19orf23 (BC338531), complete sequence.  
 ACCESSION AC005392  
 MID 93399669  
 VERSION AC005392.1 GI:3399669  
 KEYWORDS HTG.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 205251)

REFERENCE 1  
 AUTHORS Lamerdin, J.E., McCreedy, P.M., Skovronski, E., Adamson, A.W., Burhardt-Schultz, K., Gordon, L., Kyle, A., Ramirez, M., Stillwagen, S., Phan, H., Velasco, N., Do, L., Regala, W., Terry, A., Ganes, J., Dangnanan, L., Poundstone, P., Christensen, M., Georgescu, A., Ayala, J., Liu, S., Altix, C., Andreise, T., Trankiel, M., Amico-Keller, G., Coefield, J., Duarte, S., Lucas, S., Brice, R., Thomas, P., Quan, G., Krommiller, B., Arellano, A., Montgomery, M., Ow, D., Nolan, M., Trom, S., Kobayashi, A., Olsen, A.S. and Carrano, A.V.  
 TITLE Sequence analysis of a 2.5 Mb region in 19q13.2 containing a clustered CEA/PSG gene family  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 205251)  
 AUTHORS Lamerdin, J.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-1998) Joint Genome Institute, Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
 COMMENT Map and sequence oriented from q centromere to telomere. BAC 338531 (C19orf23) is adjacent to cosmid F24498 to the left and overlaps BAC 52850 (C19orf23) to the right by approximately 35 kb. Accession does not represent the sequence of the entire BAC clone insert at this time. Additional chr 19 map and sequence information are available at:  
 http://www.bio.livl.gov/genome/genome.html.  
 location/Qualifiers  
 1..205251  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="49023"  
 /chromosome="19"  
 /map="19q13.2 between BCKDHA and D19S217"  
 /clone\_lib="C19orf23"  
 /tissue\_type="sperm"  
 /note="L1NF clone name: BC338531"  
 /complement(566..1150)  
 /rpt\_family="MER67C"  
 repeat\_region 1154..1359  
 repeat\_region

[illegible]

```

/Note="BLASTN similarity to t39371 (47.77%); match: 0.96,
score: 1.7e-95; database searched: est; ya05b01.r1 Homo
sapiens cDNA clone 60553 5' similar to SP:A33514 A33514
PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN PRECURSOR --"
complement(24162..24437)
/Note="BLASTX similarity to (238..329); match: 0.9,
misc_feature
Query Match      13.2%; Score 167.4; DB 11; Length 205251;
Best Local Similarity 81.3%; Pred. No. 3.9e-29;
Matches 244; Conservative 0; Mismatches 44; Indels 12; Gaps 4;

OY 812 CCGAGCCCTCTCTCTCAGAGCCAGAGTCCAGCCGCCCTCTCTCTCAAC 871
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201698 CTCAGCCCTCTCTCTCAGAGCCAGAGTCCAGCCGCCCTCTCTCAAC 201639

OY 872 AAGG-ATACAGATCCAGAGCCCTCTCTCTCAGAGCCAGAGTCCAGAGCC 930
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201638 CAGAGATCCAGAGCCCTCTCTCTCTCTCAGAGCCAGAGTCCAGAG 201580

OY 931 CTCTCCCTCAGAGCCAGAGT-----CCAGGCCCTCTCTCTCTCTCTCT 982
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201579 CTCTCTCTCTCAGAGCCAGAGTCCAGAGCCGCCCTCTCTCTCTCTCT 201520

OY 983 CAGAGCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1042
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201519 CAG--GGCCCAAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 201462

OY 1043 TCAGATCAGAGTCCAGAGCCGCCCTCTCTCTCTCTCTCTCTCTCTCT 1102
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 201461 TCCAGCCAGAGTCCAGAGCCGCCCTCTCTCTCTCTCTCTCTCTCTCT 201402

RESULT 14
HSCPT1175/c 6592 bp DNA PRI 27-APR-1999
LOCUS Homo sapiens TNNT3 gene.
DEFINITION X90780 X90781 X90782
ACCESSION 94071059
VERSION X90780.1 GI:4071059
KEYWORDS cardiac tropoin I; TNNT3 gene.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 6592)
AUTHORS Bhavsar,P.K., Brand,N.J., Yacoub,M.H. and Barton,P.J.R.
TITLE Isolation and characterization of the human cardiac tropoin I gene
(JNNI3)
JOURNAL Genomics 35 (1), 11-23 (1996)
MEDLINE 96299735
REFERENCE 2 (bases 1 to 6592)
AUTHORS Barton,P.J.R., Cullen,M.E., Townsend,P.J., Brand,N.J., Mullen,A.J.,
Norman,D.A.M., Bhavsar,P.K. and Yacoub,M.H.
TITLE Close physical linkage of human tropoin genes: organization,
sequence, and expression of the locus encoding cardiac tropoin I
and slow skeletal tropoin I
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 6592)
AUTHORS Bhavsar,P.K.
TITLE Direct Submission
JOURNAL Submitted (15-AUG-1995) P.K. Bhavsar, Cardiothoracic Surgery,
National Heart & Lung Institute, Imperial College of Science,Tech.&
Med., Dovehouse Street, London SW3 6LJ, UK
REMARK Med., Dovehouse Street, London SW3 6LJ, UK
REFERENCE 4 (bases 1 to 6592)
AUTHORS Barton,P.J.R.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1998) Barton P.J.R., Cardiothoracic Surgery,
National Heart & Lung Institute, Imperial College of Science,Tech.&
Med., Dovehouse Street, London SW3 6LJ, UK
COMMENT On Dec 29, 1998 this sequence version replaced gi:1524065.
This sequence is directly followed by the slow skeletal tropoin I
gene (AC A7011712, A7011713)

```

```

FEATURES             Related sequence M64247.
Source               Location/Qualifiers
                     1..6592
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone_lib="P1"
                     /clone="DMPC-HEF#1-1495-E10"
                     /map="p13.3-q13.3"
                     1075..1081
                     /note="putative"
                     join(1107..1260,1532..1544,1706..1789,2195..2236,
                     2507..2638,4009..4098,4633..4809,6442..6592)
                     /gene="TNNT3"
                     1107..6592
                     /gene="TNNT3"
                     1107..1249
                     /gene="TNNT3"
                     /number=1
                     prtm_transcript 1107..6592
                     /gene="TNNT3"
                     1107..1249
                     /gene="TNNT3"
                     CDS
                     join(1250..1260,1532..1544,1706..1789,2195..2236,
                     2507..2638,4009..4098,4633..4809,6442..6525)
                     /gene="TNNT3"
                     /codon_start=1
                     /product="cardiac tropoin I"
                     /protein_id="CA62301.1"
                     /db_xref="PID:e196058"
                     /db_xref="PID:g1524066"
                     /db_xref="GI:1524066"
                     /db_xref="SWISS-PROT:P19429"
                     /translation="MADGSSDARERPPAPPIRRSSVYRAYATPEPHAKKSIAS
                     RKIQKTLQIKQELERAEERREKRALSTRCQPLAGLAEADLCROLHA
                     RKVDVEERDIEAKVTKNITEADLTOKIFDLRGFKRPTLRVRSIDAMQALLG
                     ARAKESLDLRAHLKQYKKEDTERENREVGDMRKNIIDALSGMEKRRKKEES"
                     1532..1544
                     /gene="TNNT3"
                     /number=2
                     1706..1789
                     /gene="TNNT3"
                     /number=3
                     2195..2236
                     /gene="TNNT3"
                     /number=4
                     2507..2638
                     /gene="TNNT3"
                     /number=5
                     4009..4098
                     /gene="TNNT3"
                     /number=6
                     4633..4809
                     /gene="TNNT3"
                     /number=7
                     6442..6592
                     /gene="TNNT3"
                     /number=8
                     6526..6592
                     /gene="TNNT3"
                     6568..6573
                     /gene="TNNT3"
                     /note="putative"
                     BASE COUNT 1424 a 1836 c 2048 g 1284 t
                     ORIGIN
Query Match      13.1%; Score 165.4; DB 10; Length 6592;
Best Local Similarity 70.6%; Pred. No. 1.3e-28;
Matches 262; Conservative 0; Mismatches 97; Indels 12; Gaps 3;
| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```
DB 765 CAAGAGTCCAGGCCCCAGCCCTCTCTCCCTCAGACCCAGAGTCCAGGCCCCAGCTCC 706
QY 858 TCCCTCCCTCAAAAGGAGGTACAGATCCAGGCCCCCTCCCTCAGACCCAGGAGTCCA 917
DB 705 TCTTCCCTCAGACCCAGAGTCCAGGCCCCCTCTCTCCCTCAGACCCAGAGTCCA 646
QY 918 GACCCCCCAAGCCCTCTCTCTCCCTCAGACCCAGAGT-----CAGCCCTCTCTCTCT 969
DB 645 G--GCCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGGCCCCCAGCCCTCTCTCT 588
QY 970 AGACCCAGAGTCCAGACCCCCAGCCCTCTCTCTCTCAGACCCAGGAGGTTGAGCCCC 1029
DB 587 AGACCCAGAGGCTCAGGCCCCCAGCCCTCTCTCAGACCCAGGAGTCCAGGAGCCCC 528
QY 1030 A--ACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1087
DB 527 AGCAGACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 468
QY 1088 AGAGGTNNAGTCCAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1147
DB 467 GGAGTCCAGAGCCCAATCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 408
QY 1148 TCCCTGNACAC 1158
DB 407 TCCCTCTCAGAC 397

RESULT 15
AB006136 4033 bp DNA PRI 20-JUN-1998
LOCUS Homo sapiens gene for: alpha 1,2-fucosyltransferase, 5' flanking
DEFINITION region and partial cds.
ACCESSION AB006136
NID 93242453
AB006136.1 GI:3242453
VERSION FUT1; alpha (1, 2) fucosyltransferase.
KEYWORDS Homo sapiens peripheral leukocytes DNA.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 4033)
REFERENCE 1
AUTHORS Koda, Y.
TITLE Direct Submision
JOURNAL Submitted (04-AUG-1997) to the DDBJ/EMBL/Genbank databases. Yoshiro
Koda, Kurume University, School of Medicine, Department of Forensic
Medicine, Asahimachi 67, Kurume, Fukuoka 830, Japan
(E-mail: ykoda@med.kurume-u.ac.jp, Tel:0942-31-7554,
Fax:0942-31-7700)
2 (sites)
REFERENCE 2
AUTHORS Koda, Y., Soejima, M. and Kimura, H.
TITLE Changing transcription start sites in H type
a(1,2)fucosyltransferase gene (FUT1) during differentiation of
human Erythroid lineage
JOURNAL Unpublished (1998)
COMMENT On Jun 20, 1998 this sequence version replaced gi:2317262.
Sequence updated (16-Jun-1998).
FEATURES
source
1.4033
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/map="19q13.3"
/tissue_type="peripheral leukocytes"
1..510
promoter
511..578
exon
511..578
/gene="FUT1"
gene
511..578
/gene="FUT1"
CDS
511..578
/gene="FUT1"
/standard_name="alpha(1,2)fucosyltransferase"
/codon_start=1
```

```
/product="FUT1"
/protein_id="BAA28952.1"
/db_xref="PID:d1029892"
/db_xref="PID:g3242454"
/db_xref="GI:3242454"
/translation="MGFHHVGOAGLELTGSDLPAMW"
BASE COUNT 877 a 1242 c 1099 g 815 t
ORIGIN

Query Match 12.9%; Score 163.8; DB 9; Length 4033;
Best Local Similarity 83.7%; Pred. No. 3.2e-28;
Matches 236; Conservative 0; Mismatches 33; Indels 13; Gaps 4;

QY 817 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 875
DB 2657 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2716
QY 876 GTACAGATCCAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 935
DB 2717 GTCCAGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2773
QY 936 CCTCTAGACCCAGAG-----TCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 987
DB 2774 CCTCTAGACCCAGAGGCCCAAGATCCCAAGCCCTCTCTCTCTCTCTCTCTCTCT 2833
QY 988 CCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1047
DB 2833 TCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2892
QY 1048 GTCCAGAGTCCAGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1089
DB 2893 CCGAGGATCCAGACCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2934

Search completed: September 25, 1999, 23:25:55
Job time: 6698 sec
```



---

**This Page Blank (uspto)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 25, 1999, 09:56:03 ; Search time 273 Seconds  
(without alignments)  
1159.315 Million cell updates/sec

Title: US-09-030-606-173  
Perfect score: 1265  
Sequence: 1 GGCAGCCCGCAGCTCGACGCC.....AGAGANGCAAAAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database: N\_Geneseq\_36:\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1254	99.1	1265	1	V58645
2	1254	99.1	1265	1	V61250
3	1042.4	82.4	1248	1	V58644
4	1042.4	82.4	1248	1	V61249
5	1036.2	81.9	1167	1	V58647
6	1036.2	81.9	1167	1	V61252
7	871.6	68.9	1386	1	V11855
8	641	50.7	871	1	V37495
9	474	37.5	1459	1	V58646
10	474	37.5	1459	1	V61251
11	361.4	28.6	1119	1	V58648
12	361.4	28.6	1119	1	V61253
13	346.8	27.4	402	1	X41114
14	184.6	14.6	8174	1	O13332
15	184.6	14.6	8174	1	O55908
16	184.6	14.6	8174	1	T61677
17	154	12.2	234	1	V58522
18	154	12.2	234	1	V61168
19	143	11.3	4067	1	O63192
20	118.4	9.4	1476	1	X16295
21	109.8	8.7	1438	1	T79126
22	109.8	8.7	1526	1	V07152
23	106.2	8.4	732	1	O53487
24	96.6	7.6	1089	1	T39783
25	95	7.5	986	1	O81203
26	87.4	6.9	90	1	V58593
27	87.4	6.9	90	1	V61208
28	87.2	6.9	1146	1	V84589
29	86.8	6.9	833	1	V42925
30	85.8	6.8	693	1	V61861
31	81	6.4	897	1	T49878
32	76.4	6.0	1003	1	T13316
33	75.4	6.0	741	1	N81633
34	75.4	6.0	744	1	T03999
35	75.4	6.0	744	1	T04000
36	75.4	6.0	744	1	T04001
37	75.4	6.0	790	1	V24548
38	73.8	5.8	699	1	V16367
39	71.4	5.6	683	1	O63794
40	70.8	5.6	907	1	N70905
41	70.4	5.6	925	1	N93196
42	69.8	5.5	701	1	O63795
43	69.4	5.5	701	1	X23298

## ALIGNMENTS

RESULT 1  
ID V58645  
V58645 standard; cDNA: 1265 BP.  
AC V58645;  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE2.  
KW Prostate tumour specific gene; human; prostate cancer; detection; therapy; ss.  
OS Homo sapiens.  
PN W09837418-A2.  
PD 27-AUG-1998.  
PR 25-FEB-1998: U03690.  
PR 09-FEB-1998: US-904809.  
PR 25-FEB-1997: US-806596.  
PR 01-AUG-1997: US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI: 98-480805/41.  
PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers  
PS Claim 1; Page 113-114; 14pp; English.  
CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprising contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.  
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match 99.1%; Score 1254; DB 1; Length 1265;  
Best Local Similarity 100.0%; Pred. No. 3.9e-296;  
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGCAGCCCGCAGCTCGACGCCCTGGAGCGGCAGCTGTCATGGAACGAATTGTTTCG 60  
DB 1 GGCAGCCCGCAGCTCGACGCCCTGGAGCGGCAGCTGTCATGGAACGAATTGTTTCG 60  
QY 61 TCGGGGCTCTGTCATCGCAGTGGGGTGTGTCAGCCGACAGCTTTCCAGAACTCC 120  
DB 61 TCGGGGCTCTGTCATCGCAGTGGGGTGTGTCAGCCGACAGCTTTCCAGAACTCC 120  
QY 121 TACACATCGGGCTGGGCTGACAGTCTTGGAGCGGACCAAGAGAGAGAGAGATG 180  
DB 121 TACACATCGGGCTGGGCTGACAGTCTTGGAGCGGACCAAGAGAGAGAGAGATG 180  
QY 121 TACACATCGGGCTGGGCTGACAGTCTTGGAGCGGACCAAGAGAGAGAGAGATG 180  
DB 121 TACACATCGGGCTGGGCTGACAGTCTTGGAGCGGACCAAGAGAGAGAGAGATG 180  
QY 181 GTGAGAGCCAGCTCTCGTACGCAACCCAGTACACAGACCTTGTCTTAACGAC 240  
DB 181 GTGAGAGCCAGCTCTCGTACGCAACCCAGTACACAGACCTTGTCTTAACGAC 240  
QY 181 GTGAGAGCCAGCTCTCGTACGCAACCCAGTACACAGACCTTGTCTTAACGAC 240  
DB 181 GTGAGAGCCAGCTCTCGTACGCAACCCAGTACACAGACCTTGTCTTAACGAC 240  
QY 241 CTCATGCTCATGAATGAGTGAATCCGTCGAGTGTGACACCATCCGGAGCATCAC 300  
DB 241 CTCATGCTCATGAATGAGTGAATCCGTCGAGTGTGACACCATCCGGAGCATCAC 300  
QY 301 ATTGCTTCAGTGCCTTACCGCGGGGAACCTTTCCTGCTGGGTGCTGTG 360  
DB 301 ATTGCTTCAGTGCCTTACCGCGGGGAACCTTTCCTGCTGGGTGCTGTG 360  
QY 361 GCGAAGGTTGACCTACGGGTGTGTCTGCTTCAAGAGAGTCTTCCCAATCG 420  
DB 361 GCGAAGGTTGACCTACGGGTGTGTCTGCTTCAAGAGAGTCTTCCCAATCG 420  
QY 421 CGGGGGCTACCCAGAGCTCTGCTGCCAGGAGATGCTACCTGCTGCGAGTGGTGA 480  
DB 421 CGGGGGCTACCCAGAGCTCTGCTGCCAGGAGATGCTACCTGCTGCGAGTGGTGA 480

```

Db      |||||||
421     GGGGGGCTGACCCAGAGCTGCTGCTCCAGCGAGATGCTCTACCTGCTGAGTGCATGA 480
Qy      |||||||
481     ACGTGTGGTGTCTGTAGAGAGTCTGTAGTAACTGATGACCGGTATCCACCCCA 540
Db      |||||||
481     ACGTGTGGTGTCTGTAGAGAGTCTGTAGTAACTGATGACCGGTATCCACCCCA 540
Qy      |||||||
541     GCATGTCTGCGCGCGGAGGAGGCAAGACAGAGAGTCTGTGCAACGGTACTGTGGGG 600
Db      |||||||
541     GCATGTCTGCGCGCGGAGGAGGCAAGACAGAGAGTCTGTGCAACGGTACTGTGGGG 600
Qy      |||||||
601     GGGCCCTATCTGCAACGGGTACTGTGCAAGGGCTTGTCTTTCGGAAGACCCGTGTG 660
Db      |||||||
601     GGGCCCTATCTGCAACGGGTACTGTGCAAGGGCTTGTCTTTCGGAAGACCCGTGTG 660
Qy      |||||||
661     GCCAAGTTGGCGTGTCTGTACCAACCTGTGCAAAATTCATGATGATAGAGA 720
Db      |||||||
661     GCCAAGTTGGCGTGTCTGTACCAACCTGTGCAAAATTCATGATGATAGAGA 720
Qy      |||||||
721     AAACCGTCCAGGCGCAATTAACTCTGGGACTGGGAACCATGAATTCACCCCAATAC 780
Db      |||||||
721     AAACCGTCCAGGCGCAATTAACTCTGGGACTGGGAACCATGAATTCACCCCAATAC 780
Qy      |||||||
781     ATCTGCGGAAGGAATTACGAATATCTGTTCCAGCCCTCTCTCCAGCCGAGAG 840
Db      |||||||
781     ATCTGCGGAAGGAATTACGAATATCTGTTCCAGCCCTCTCTCCAGCCGAGAG 840
Qy      |||||||
841     TCCAGAGCCCGAGCCCTCTCTCTCAAAACCAAGGTAAGATCCCGAGCCCTCTCTCC 900
Db      |||||||
841     TCCAGAGCCCGAGCCCTCTCTCTCAAAACCAAGGTAAGATCCCGAGCCCTCTCTCC 900
Qy      |||||||
901     TCAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCTCAGACCCAGAGAGTCCAGCCCC 960
Db      |||||||
901     TCAGACCCAGAGTCCAGACCCCGAGCCCTCTCTCTCAGACCCAGAGAGTCCAGCCCC 960
Qy      |||||||
961     TCCTCTCAGACCCAGAGAGTCCAGACCCCGAGCCCTCTCTCTCAGACCCAGAGGTT 1020
Db      |||||||
961     TCCTCTCAGACCCAGAGAGTCCAGACCCCGAGCCCTCTCTCTCAGACCCAGAGGTT 1020
Qy      |||||||
1021     GAGGCGCCCAACCCCTCTCTCTCAGAGTCCAGAGGTTCCAGCCCGCAACCCCTGTTCC 1080
Db      |||||||
1021     GAGGCGCCCAACCCCTCTCTCTCAGAGTCCAGAGGTTCCAGCCCGCAACCCCTGTTCC 1080
Qy      |||||||
1081     CAGACCCAGAGGTTNAGGTCCAGCCCTCTCTCTCAGACCCAGAGGTTCCATGCCACC 1140
Db      |||||||
1081     CAGACCCAGAGGTTNAGGTCCAGCCCTCTCTCTCAGACCCAGAGGTTCCATGCCACC 1140
Qy      |||||||
1141     TAGATTTTCCCTGACACAGTGCCTCTGTTGAGNAGTTGACCCCAACTTACCACTTGGT 1200
Db      |||||||
1141     TAGATTTTCCCTGACACAGTGCCTCTGTTGAGNAGTTGACCCCAACTTACCACTTGGT 1200
Qy      |||||||
1201     TTTTCATTTTNGTCCCTTCCCTAGATCCGAATTAAGTTTAAGAGANGCAAAAA 1260
Db      |||||||
1201     TTTTCATTTTNGTCCCTTCCCTAGATCCGAATTAAGTTTAAGAGANGCAAAAA 1260
Qy      |||||||
1261     AAAAA 1265
Db      |||||||
1261     AAAAA 1265

```

```

PR      01-AUG-1997; US-904804.
PA      (CORI-) CORIXA CORP.
PI      Dillon DC, Xu J.
PI      WPI: 98-609886/51.
PT      Polypeptides comprising immunogenic portions of prostate proteins -
/PT      used in a vaccine for the treatment of prostate cancer
/CC      Claim 3; Page 105-106; 130pp; English.
/CC      The present sequence is a new DNA which encodes an immunogenic portion
/CC      of a prostate tumour protein. The encoded immunogen, or the DNA itself,
/CC      can be used as a vaccine for the treatment of prostate cancer. The DNA
/CC      was identified by analysis of a subtracted cDNA library obtained by
/CC      subtracting a prostate tumour cDNA expression library with a normal
/CC      tissue cDNA library.
SQ      Sequence. 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match      99.18; Score 1254; DB 1; Length 1265;
Best Local Similarity 100.0%; Pred. No. 3,9e-296;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 GGCAGCCCGCACTGGCAGGCCCTGGCAGGCGCACTGGTCAATGAAAAAGATTGTTGCG 60
Db      1 GGCAGCCCGCACTGGCAGGCCCTGGCAGGCGCACTGGTCAATGAAAAAGATTGTTGCG 60
Qy      61 TCGGGCGTCTGTGTCATCCGACAGTGGGTGCTGTACAGCCGACACTGTCTTCCAGAACTCC 120
Db      61 TCGGGCGTCTGTGTCATCCGACAGTGGGTGCTGTACAGCCGACACTGTCTTCCAGAACTCC 120
Qy      121 TACACCATGCGGGCTGGGCTCTGACAGTCTTGAGGCCGACCAAGACCAAGGAGCCAGATG 180
Db      121 TACACCATGCGGGCTGGGCTCTGACAGTCTTGAGGCCGACCAAGACCAAGGAGCCAGATG 180
Qy      181 GTGGAGGCGACCTCTCTCGTACGGGACCCAGAGTACAAAGACCTGCTCGTAAAGAC 240
Db      181 GTGGAGGCGACCTCTCTCGTACGGGACCCAGAGTACAAAGACCTGCTCGTAAAGAC 240
Qy      241 CTCAATGCTCATCAAGTTGAGCAATCCGTGTCCAGTCTGACACATCCGAGAGCATCAGC 300
Db      241 CTCAATGCTCATCAAGTTGAGCAATCCGTGTCCAGTCTGACACATCCGAGAGCATCAGC 300
Qy      301 ATTGCTTGCAGTGCCTTACCGCGGGGAACCTTGTGCTTGTGCTGGGCTCTCTG 360
Db      301 ATTGCTTGCAGTGCCTTACCGCGGGGAACCTTGTGCTTGTGCTGGGCTCTCTG 360
Qy      361 GCGAAGGTGAGTCCAGGGGTGTGTGCTGCTCTTCAAGAGAGTCTCTGCCAATCG 420
Db      361 GCGAAGGTGAGTCCAGGGGTGTGTGCTGCTCTTCAAGAGAGTCTCTGCCAATCG 420
Qy      421 CGGGGGCTGACCCAGAGTCTGCTGCCAGGACAGAAATGCTTACCGTCTGCAATGCGTGA 480
Db      421 CGGGGGCTGACCCAGAGTCTGCTGCCAGGACAGAAATGCTTACCGTCTGCAATGCGTGA 480
Qy      481 ACGTGTGGTGTGTCTGAGGAGTCTGCACTAAGTCTATGACCCGCTGACACCCCA 540
Db      481 ACGTGTGGTGTGTCTGAGGAGTCTGCACTAAGTCTATGACCCGCTGACACCCCA 540
Qy      541 GCATGTCTGCGCGGAGGAGGCAAGACCAAGAGACTCTGCAAGGTAATCTCTGGG 600
Db      541 GCATGTCTGCGCGGAGGAGGCAAGACCAAGAGACTCTGCAAGGTAATCTCTGGG 600
Qy      601 GGGCCCTATCTGCAACGGGTACTGTGCAAGGGCTTGTCTTTCGGAAGACCCGTGTG 660
Db      601 GGGCCCTATCTGCAACGGGTACTGTGCAAGGGCTTGTCTTTCGGAAGACCCGTGTG 660
Qy      661 GCCAAGTTGGCGTGTCTGTACCAACCTGTGCAAAATTCATGATGATAGAGA 720
Db      661 GCCAAGTTGGCGTGTCTGTACCAACCTGTGCAAAATTCATGATGATAGAGA 720
Qy      721 AAACCGTCCAGGCGCAATTAACTCTGGGACTGGGAACCATGAATTCACCCCAATAC 780
Db      721 AAACCGTCCAGGCGCAATTAACTCTGGGACTGGGAACCATGAATTCACCCCAATAC 780
Qy      781 ATCTGCGGAAGGAATTACGAATATCTGTTCCAGCCCTCTCTCTCAGAGCCGAGAG 840

```

```

      |||||||
Db 781 ATCTGGGAAGGAATTCAGAAATATCTGTCCAGGCCCTCTCCCTCAGGCCAGAG 840
      |||||||
Oy 841 TCCAGGCCCCCAGCCCTCTCTCTCTCAAAACCAAGGTAACAGATCCCCAGCCCTCTCC 900
      |||||||
Db 841 TCCAGGCCCCCAGCCCTCTCTCTCTCAAAACCAAGGTAACAGATCCCCAGCCCTCTCC 900
      |||||||
Oy 901 TCAGACCAGAGATTCAGACACCCCGAGCCCTCTCTCTCTAGACCCAGGAGTCCAGCC 960
      |||||||
Db 901 TCAGACCAGAGATTCAGACACCCCGAGCCCTCTCTCTCTAGACCCAGGAGTCCAGCC 960
      |||||||
Oy 961 TCTCTCTCAGACCCAGAGTTCAGACACCCCGAGCCCTCTCTCTCTAGACCCAGGAGT 1020
      |||||||
Db 961 TCTCTCTCAGACCCAGAGTTCAGACACCCCGAGCCCTCTCTCTCTAGACCCAGGAGT 1020
      |||||||
Oy 1021 GAGGCCCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
      |||||||
Db 1021 GAGGCCCCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
      |||||||
Oy 1081 CAGACCCAGAGTNNAGTCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
      |||||||
Db 1081 CAGACCCAGAGTNNAGTCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1140
      |||||||
Oy 1141 TAGATTTTCCCTGNACACAGTGCCTCTGTGTGNANGTGACCACTTACAGTTGGT 1200
      |||||||
Db 1141 TAGATTTTCCCTGNACACAGTGCCTCTGTGTGNANGTGACCACTTACAGTTGGT 1200
      |||||||
Oy 1201 TTTTCATTTTNGTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
      |||||||
Db 1201 TTTTCATTTTNGTCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
      |||||||
Oy 1261 AAAAA 1265
      |||||||
Db 1261 AAAAA 1265

```

## RESULT 3

VS8644  
ID VS8644 standard; cDNA: 1248 BP.

AC VS8644;

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DEL.

KM Prostate tumour specific gene; human; prostate cancer; detection;

OS Homo sapiens.

FN key Location/Qualifiers

FT CDS 217..696

FT CDS /\*tag- a

PN MO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998: U01690.

PR 09-FEB-1998: US-904809.

PR 25-FEB-1997: US-806596.

PR 01-AUG-1997: US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J.

DR WPI: 98-480805/41.

DR P-PSD: W69387.

PT Novel human prostate specific tumour protein and fragments - useful

PS Claim 1; Page 112; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can

be used in the method of the invention. The method is for detecting

prostate cancer comprises contacting a biological sample with an agent

able to bind an immunogenic portion of a prostate protein (such as

encoded by this sequence). An antibody which binds to an immunogenic

portion of the prostate protein, and the method can be used to detect,

monitor progression of, or treat prostate cancer. The antibody may

also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;

Query Match 82.4%; Score 1042.4; DB 1; Length 1248;  
Best Local Similarity 91.1%; Pred. No. 1,1e-244;  
Matches 1163; Conservative 0; Mismatches 18; Indels 95; Gaps 2;

```

      |||||||
Oy 2 GCAGCCCGCAGCTGCAGCCCTCTGAGGCGGCGACTGTGATGAAACGAATGTTCTGCT 61
      |||||||
Db 29 GCAGCCCGCAGCTGCAGCCCTCTGAGGCGGCGACTGTGATGAAACGAATGTTCTGCT 88
      |||||||
Oy 62 CGGGGCTCTCTGTGATCCGAGTGGGTGCTGTACAGCCGACACTGTTTCCAGAA----- 117
      |||||||
Db 89 CGGGGCTCTCTGTGATCCGAGTGGGTGCTGTACAGCCGACACTGTTTCCAGAAAGTAG 148
      |||||||
Oy 117 -----CTCTCACCATTGGGGGTGGGCTGTGACAGTCTTGGAGCCGACCAAGAGCAG 169
      |||||||
Db 149 TGCAGAGCTCTTACACCATATGGGCTGGGCTGTGACAGTCTTGGAGCCGACCAAGAGCAG 208
      |||||||
Oy 170 GAGCCAGATGTGTGAGGCGGAGCTCTCGTAGGCGACCCAGAGTACAGACACCTTGC 229
      |||||||
Db 209 GAGCCAGATGTGTGAGGCGGAGCTCTCGTAGGCGACCCAGAGTACAGACACCTTGC 268
      |||||||
Oy 230 TCGCTAACAGACCTCATCTATCATAGTTGACGAATCCGTGTCCGATCTGACACCATCC 289
      |||||||
Db 269 TCGCTAACAGACCTCATCTATCATAGTTGACGAATCCGTGTCCGATCTGACACCATCC 328
      |||||||
Oy 290 GAGCATCAGCATGTTCTTCCAGTGCCTACCGGGGGGAATCTTGGCTGTGCTTGGCT 349
      |||||||
Db 329 GAGCATCAGCATGTTCTTCCAGTGCCTACCGGGGGGAATCTTGGCTGTGCTTGGCT 388
      |||||||
Oy 350 GGGGTCTGTGTGGCGAACGGGTGAGTCAAGGGGTGTGTGTCTGCTCTTCAAGAGGTCT 409
      |||||||
Db 389 GGGGTCTGTGTGGCGAAC-----GGCAGAAATGCTTACCGTGTCT 406
      |||||||
Oy 410 CTGCCCAGTGGGGGGCTGACCCAGACTCTGCTCTCCAGGCAATGCTTACCGTGTCT 469
      |||||||
Db 406 -----GGCAGAAATGCTTACCGTGTCT 425
      |||||||
Oy 470 GCAGTGCGTGAAGTGTGCGGTGTGTGAGAGAGTGTGCAAGTATGATACCGGCT 529
      |||||||
Db 426 GCAGTGCGTGAAGTGTGCGGTGTGTGAGAGAGTGTGCAAGTATGATACCGGCT 485
      |||||||
Oy 530 GTACACCCCGCAGCATGTTCTTGGCGGCGGAGGCGCAAGACAGAGACTCTGCAACGG 589
      |||||||
Db 486 GTACACCCCGCAGCATGTTCTTGGCGGCGGAGGCGCAAGACAGAGACTCTGCAACGG 545
      |||||||
Oy 590 TGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTCCAGGGCTTGTCTTTCGAAA 649
      |||||||
Db 546 TGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTCCAGGGCTTGTCTTTCGAAA 605
      |||||||
Oy 650 AGCCCGGTGTGGCCAGTGTGGCGGTGCGAGGTGTACACCAACCTGCAAAATCACTGA 709
      |||||||
Db 606 AGCCCGGTGTGGCCAGTGTGGCGGTGCGAGGTGTGTACCAACCTGCAAAATCACTGA 665
      |||||||
Oy 710 GTGATAGAGAAAACCGTCCAGGCGCACTTACTGTGGGAGTGGGAACCCATGAATGA 769
      |||||||
Db 666 GTGATAGAGAAAACCGTCCAGGCGCACTTACTGTGGGAGTGGGAACCCATGAATGA 725
      |||||||
Oy 770 CCCCCAATATATCTCTCGGAGGAATTCAGGAATATCTGTCCAGCCCTCTCTCTC 829
      |||||||
Db 726 CCCCCAATATATCTCTCGGAGGAATTCAGGAATATCTGTCCAGCCCTCTCTCTC 785
      |||||||
Oy 830 AGGCGGAGAGTTCAGGCGCCCGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 889
      |||||||
Db 786 AGGCGGAGAGTTCAGGCGCCCGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 845
      |||||||
Oy 890 CCCCCTCTCTCTAGACCCAGAGTCCAGACCCCGACCCCTCTCTCTCTCTCTCTCTCTCT 949
      |||||||
Db 846 CCCCCTCTCTCTAGACCCAGAGTCCAGACCCCGACCCCTCTCTCTCTCTCTCTCTCTCT 905
      |||||||
Oy 950 AGTCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1009
      |||||||
Db 906 AGTCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 965
      |||||||
Oy 1010 ACCCAGGGGTTGAGGCGCCCGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1069

```

|||||  
Db 966 ACCCAGGGGTCCAGGGCCCCCAACCCCTCCCTCAGACTCAGAGGTCCAGGCCCAAC 1025  
Oy 1070 CCTCGTTCCCAAGACCCAGAGTNNAGTCCAGGCCCTCTCTCCATCAGACCCAGNGT 1129  
Db 1026 CCTTCATTCCTCCAGAGCCAGAGTCCAGGCTCCAGCCCTCCTCCTCAGACCCAGGT 1085  
Oy 1130 CCAATGCCACCTAGATTTTCCCTGNACAGAGTCCCTTGTGNNGTGACCAACCT 1189  
Db 1086 CCAATGCCACCTAGACTNTCCCTGACAGAGTCCCTCTTGTGACAGCTTGAACCACT 1145  
Oy 1190 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1249  
Db 1146 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1205  
Oy 1250 AANGCAAAAAAAA 1265  
Db 1206 AANGCAAAAAAAA 1221

RESULT 4  
V61249  
ID V61249 standard; cDNA: 1248 BP.  
AC V61249;  
DT 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DE1.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN M09837093-R2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; 003492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI D110n DC, Xu J;  
DR MPI: 98-609886/51.  
DR P-PsDB: W71871.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PS used in a vaccine for the treatment of prostate cancer  
PS Claim 3; Page 104; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
CC Sequence 1248 BP; 288 A; 424 C; 303 G; 228 T;  
SQ

Query Match 82.4%; Score 1042.4; DB 1; Length 1248;  
Best Local Similarity 91.1%; Pred. No. 1,1e-244;  
Matches 1163; Conservative 0; Mismatches 18; Indels 95; Gaps 2;

Oy 290 GGAGATCAGACTGCTTGCGAGTGCCTACCGGGGAACTTGGCTTCTGGCT 349  
Db 329 GGAGATCAGACTGCTTCTTGCGAGTGCCTACCGGGGAACTTGGCTTCTGGCT 388  
Oy 350 GGGGCTGCTGGCGAAGGTGAGTACAGGGGTGTGTCTGCTCCTCTTCAAGAGAGTCT 409  
Db 389 GGGGCTGCTGGCGAAG-----GGCAAAATGCTTACCGTCT 406  
Oy 410 CTGCCAGTCCGGGGGCTGACCCAGAGCTTGCCTCCAGGAGAAATGCTTACCGTCT 469  
Db 406 -----GGCAAAATGCTTACCGTCT 425  
Oy 470 GCAGTGCCTGAACGTGCGGTGTGCTGAGAGGTGCTGCAATGCTTATGACCGCT 529  
Db 426 GCAGTGCCTGAACGTGCGGTGTGCTGAGAGGTGCTGCAATGCTTATGACCGCT 485  
Oy 530 GTACACCCACAGATGTTCTGCGCGCGGAGGCAAGACAGAGACTCTGCAACGG 589  
Db 486 GTACACCCACAGATGTTCTGCGCGCGGAGGCAAGACAGAGACTCTGCAACGG 545  
Oy 590 TGACTCTGGGGGGCCCTGATCTGCAAGGGTACTTGCAGGGCTTGTGTCTTTGGAAA 649  
Db 546 TGACTCTGGGGGGCCCTGATCTGCAAGGGTACTTGCAGGGCTTGTGTCTTTGGAAA 605  
Oy 650 ACCCCGTTGSCCAAGTTGGCGTGCAGGTGCTACACCAACCTCTGCAATTCATGA 709  
Db 606 ACCCCGTTGSCCAAGTTGGCGTGCAGGTGCTACACCAACCTCTGCAATTCATGA 665  
Oy 710 GTGATAGAGAAACCGTCCAGGCGAGTTAATCTGGGGAGTGGGAACCATGAATTGA 769  
Db 666 GTGATAGAGAAACCGTCCAGGCGAGTTAATCTGGGGAGTGGGAACCATGAATTGA 725  
Oy 770 CCCCCAAATACATCTCTGGGGAAGAAATTCAGAAATATCTGTTCCAGCCCTCTCTC 829  
Db 726 CCCCCAAATACATCTCTGGGGAAGAAATTCAGAAATATCTGTTCCAGCCCTCTCTC 785  
Oy 830 AGGCCAGAGAGTCAAGGCCCCAGCCCTCCTCCTCAACCAAGGATACAGATCCAG 889  
Db 786 AGGCCAGAGAGTCAAGGCCCCAGCCCTCCTCCTCAACCAAGGATACAGATCCAG 845  
Oy 890 CCCCTCTCTCCTCAGACCCAGAGTCCAGAGCCCCAGCCCTCTCTCTCAGACCCAG 949  
Db 846 CCCCTCTCTCCTCAGACCCAGAGTCCAGAGCCCCAGCCCTCTCTCTCAGACCCAG 905  
Oy 950 ACTCCAGCCCTCTCTCTCAGACCCAGAGTCCAGAGCCCCAGCCCTCTCTCTAG 1009  
Db 906 ACTCCAGCCCTCTCTCTCAGACCCAGAGTCCAGAGCCCCAGCCCTCTCTCTAG 965  
Oy 1010 ACCCAGGGTTGAGGCCCCCAACCCCTCTCTCTCAGATCAGAGTCCAGAGCCCAAC 1069  
Db 966 ACCCAGGGTTGAGGCCCCCAACCCCTCTCTCTCAGATCAGAGTCCAGAGCCCAAC 1025  
Oy 1070 CCTCGTTCCCAAGACCCAGAGTNNAGTCCAGGCCCTCTCTCCATCAGACCCAGNGT 1129  
Db 1026 CCAATGCCACCTAGATTTTCCCTGNACAGAGTCCCTTGTGNNGTGACCAACCT 1189  
Oy 1130 CCAATGCCACCTAGATTTTCCCTGNACAGAGTCCCTTGTGNNGTGACCAACCT 1145  
Db 1086 CCAATGCCACCTAGATTTTCCCTGNACAGAGTCCCTTGTGNNGTGACCAACCT 1145  
Oy 1190 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1249  
Db 1146 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1205  
Oy 1250 AANGCAAAAAAAA 1265  
Db 1206 AANGCAAAAAAAA 1221

RESULT 5  
V58647  
ID V58647 standard; cDNA: 1167 BP.

AC V58647;  
 DT 08-DEC-1998 (first entry)  
 DE Prostate tumour specific gene clone DEL3.  
 KM Prostate tumour specific gene; human; prostate cancer; detection;  
 OS therapy: ss.  
 HS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT CDS 28..645  
 FT /tag= a  
 PD MO9837418-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998; U03690.  
 PR 09-FEB-1998; US-904809.  
 PR 25-FEB-1997; US-806596.  
 PR 01-AUG-1997; US-904809.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI: 98-480805/41.  
 DR P-PSDB: W69388.  
 PT Novel human prostate specific tumour protein and fragments - useful  
 PT for detecting and treating prostate cancers  
 PS Claim 1: Page 115; 14pp; English.  
 CC This sequence represents a human prostate tumour specific gene, and can  
 CC be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprises contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC encoded by this sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 81.9%; Score 1036.2; DB 1; Length 1167;  
 Best Local Similarity 91.0%; Pred. No. 3,5e-243;  
 Matches 1137; Conservative 0; Mismatches 29; Indels 83; Gaps 1;

14 CGAGAGCCCTGGAGGGGCACTGGTATGAAACGAATGTTGCTGGGGGCTCTGG 73  
 Db 2 CGAGAGCCCTGGAGGGGCACTGGTATGAAACGAATGTTGCTGGGGGCTCTGG 61  
 QY 74 TGCATCCGAGTGGGTGCTGTCAGCCGACACTGTTTCCAGAACTCCTACACATCGGGC 133  
 Db 62 TGCATCCGAGTGGGTGCTGTCAGCCGACACTGTTTCCAGAACTCCTACACATCGGGC 121  
 QY 134 TGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGGGAGCCAGAGTGGAGGCCAGCC 193  
 Db 122 TGGGCTGACAGTCTTGAGGCGGACCAAGAGCCAGGGAGCCAGAGTGGAGGCCAGCC 181  
 QY 194 TCTCCGTACGGGACCCAGAGTACAAAGACCTTGTCTGCTAAGACCTATGCTATCA 253  
 Db 182 TCTCCGTACGGGACCCAGAGTACAAAGACTTGTCTGCTAAGACCTATGCTATCA 241  
 QY 254 AGTTGACGAATCGGTGCTGAGTGTGACACATCCGAGCATCAGATTGCTTCGAGT 313  
 Db 242 AGTTGACGAATCGGTGCTGAGTGTGACACATCCGAGCATCAGATTGCTTCGAGT 301  
 QY 314 GCCCTACCGGGGGAACCTTGCCTGTTTCTGGCTGGGGTCTGCTGGCAAGCGTGAGC 373  
 Db 302 GCCCTACCGGGGGAACCTTGCCTGTTTCTGGCTGGGGTCTGCTGGCAAGC----- 355  
 QY 374 TCAGGGGTGTGTCTCTCCTCTTCAAGAGAGTCTCTGCCCAATCCGGGGGCTGACCC 433  
 Db 355 ----- 355  
 QY 434 AGAGCTGTGCTCCAGGAGAAATGCTACCGTGTCTGAGTGGTGAAGCTGTGCTGTG 493  
 Db 355 -----GGCAGAAATGCTACCGTGTCTGAGTGGTGAAGCTGTGCTGTG 398  
 QY 494 GTCTGAGGAGGTCTGAGTAAGTCTATGACCCGCTGTACACCCAGCATGTTCTGCGC 553  
 Db 399 GTCTGAGGAGGTCTGAGTAAGTCTATGACCCGCTGTACACCCAGCATGTTCTGCGC 458

QY 554 CGGCGAGGGGCAAGACCAAGAGACTCTGCAACGGTGACTGTGGGGGCCCTGTATCTG 613  
 Db 459 CGGCGAGGGGCAAGACCAAGAGACTCTGCAACGGTGACTGTGGGGGCCCTGTATCTG 518  
 QY 614 CAAGGGTACTGACAGGGCCTGTGTTTGGAAAAAGCCCGGTGTGGCCAAAGTTGGCGT 673  
 Db 519 CAAGGGTACTGACAGGGCCTGTGTTTGGAAAAAGCCCGGTGTGGCCAAAGTTGGCGT 578  
 QY 674 GCCAGTGTCTACCAACCTCTGCAATTCAGTGTGATGATGAAAAACGCTCCAGGC 733  
 Db 579 GCCAGTGTCTACCAACCTCTGCAATTCAGTGTGATGATGAAAAACGCTCCAGGC 638  
 QY 734 CAGTTAACTGTGGGAGTGGGAACCATGAAATTTGACCCCAATATCATCTCTGGAAG 793  
 Db 639 CAGTTAACTGTGGGAGTGGGAACCATGAAATTTGACCCCAATATCATCTCTGGAAG 698  
 QY 794 AATTCAAGAAATATCTGTCCAGACCCCTCTCCCTCAGGCCCCAGAGTCCAGGCCCCAG 853  
 Db 699 AATTCAAGAAATATCTGTCCAGACCCCTCTCCCTCAGGCCCCAGAGTCCAGGCCCCAG 758  
 QY 854 CCCCTCTCCCTCAAAACCAAGGTTACAGATCCCAACCCCTCTCTCTCAGACCCAGAG 913  
 Db 759 CCCCTCTCCCTCAAAACCAAGGTTACAGATCCCAACCCCTCTCTCTCAGACCCAGAG 818  
 QY 914 TCCAGACCCCGGAGCCCTCTCTCAGACCCAGAGTTCAGACCCCTCTCTCAGAC 973  
 Db 819 TCCAGACCCCGGAGCCCTCTCTCAGACCCAGAGTTCAGACCCCTCTCTCAGAC 878  
 QY 974 CCAGAGTCCAGACCCCGGAGCCCTCTCTCAGACCCAGAGGTTGAGGCCCCCAAC 1033  
 Db 879 CCAGAGTCCAGACCCCGGAGCCCTCTCTCAGACCCAGAGGTTGAGGCCCCCAAC 938  
 QY 1034 CCTCTCTCTCTCAAGTTCAGAGTTCAGAGCCCAACCCCTCTCTCTCTCAGACCAAGT 1093  
 Db 939 CCTCTCTCTCTCAAGTTCAGAGTTCAGAGCCCAACCCCTCTCTCTCTCAGACCAAGT 998  
 QY 1094 NNAGGTCCCGAGCCCT 1153  
 Db 999 NCAGGTCCCGAGCCCT 1058  
 QY 1154 NCAGGTCCCGAGCCCT 1213  
 Db 1059 TACACAGTGGCCCT 1118  
 QY 1214 TCCCTTTCCCTTGAATTCAGAAATTAAGTTTAAGAGANGCAAAAAA 1262  
 Db 1119 TCCCTTTCCCTTGAATTCAGAAATTAAGTTTAAGAGANGCAAAAAA 1167

RESULT 6  
 ID V61252  
 AC V61252;  
 DT 06-JAN-1999 (first entry)  
 DE cDNA sequence of prostate tumour clone P703 splice variant DEL3.  
 KM Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
 OS Homo sapiens.  
 FT Key Location/Qualifiers  
 FT CDS 27-AUG-1998.  
 PD MO9837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998; U03492.  
 PR 09-FEB-1998; US-020956.  
 PR 25-FEB-1997; US-806099.  
 PR 01-AUG-1997; US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI: 98-609886/51.  
 DR P-PSDB: W71782.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 PT used in a vaccine for the treatment of prostate cancer  
 PS Claim 3: Page 107; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,

CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match 81.9%; Score 1036.2; DB 1; Length 1167;  
Best Local Similarity 91.0%; Pred. No. 3.5e-243;  
Matches 1137; Conservative 0; Mismatches 29; Indels 83; Gaps 1;

QY 14 CCGAGCCCTTGGCAGGCGGCGCTGTCATGTAAGAAAGATTGTTCTGCTGGGCGTCTGG 73  
DB 2 CCGAGCCCTTGGCAGGCGGCGCTGTCATGTAAGAAAGATTGTTCTGCTGGGCGTCTGG 61  
QY 74 TGCATCCGCGAGTGGTGTCTGTCAGCCGCGACACTGTTTCCAGAACTCTCTACCACTGGGC 133  
DB 62 TGCATCCGCGAGTGGTGTCTGTCAGCCGCGACACTGTTTCCAGAACTCTCTACCACTGGGC 121  
QY 134 TGGGCGTGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCGAGCC 193  
DB 122 TGGGCGTGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGTTGAGGCGAGCC 181  
QY 194 TCTCCGTACGGCAGCCAGAGTACACAGACCTTCTCTGCTTAAGCAAGCTTCACTCATCA 253  
DB 182 TCTCCGTACGGCAGCCAGAGTACACAGACTTCTCTGCTTAAGCAAGCTTCACTCATCA 241  
QY 254 AGTTGAGCAAGTCCGTGCTGCTGACACATCCGAGACATACAGATGCTTCCAGT 313  
DB 242 AGTTGAGCAAGTCCGTGCTGCTGACACATCCGAGACATACAGATGCTTCCAGT 301  
QY 314 GCCCTACCGCGGGGAACTCTTGGCTGTTTGGCTGGGGTGTCTGTCGCAACGGTGAAGC 373  
DB 302 GCCCTACCGCGGGGAACTCTTGGCTGTTTGGCTGGGGTGTCTGTCGCAACGGTGAAGC 355  
QY 374 TCACGGGTGTGTCTGCTCTTCAAGAGGTCTCTGCCACGTCGGGGGCGTGAACC 433  
DB 355 ----- 355  
QY 434 AGAGCTCTGCTGCCAGGCGAGATGCTTACCGTACAGTGGGTGAACGTGCGGTGGT 493  
DB 355 -----GCGAAGTGGCTTACCGTACAGTGGGTGAACGTGCGGTGGT 398  
QY 494 GTCTGAGAGAGTGTGAGTAACTCTATGACCGCGCTGTACACCCAGCATGTTCTGCGC 553  
DB 399 GTCTGAGAGAGTGTGAGTAACTCTATGACCGCGCTGTACACCCAGCATGTTCTGCGC 458  
QY 554 CCGCGAGAGGCGAAGACCAAGAGACTCTGCAACGGTGAAGTGGGGGCGCCCTGATCTG 613  
DB 459 CCGCGAGAGGCGAAGACCAAGAGACTCTGCAACGGTGAAGTGGGGGCGCCCTGATCTG 518  
QY 614 CACGGGTAAGTTCAGAGGCTTGTCTTTCGAAAAAGCCCGCTGTGGCAATTTGGCGT 673  
DB 519 CACGGGTAAGTTCAGAGGCTTGTCTTTCGAAAAAGCCCGCTGTGGCAATTTGGCGT 578  
QY 674 GCCAGGTGTCTACCAACCTCTGCAAAATTCAGTGAAGTGAAGAAACCTGCGAGG 733  
DB 579 GCCAGGTGTCTACCAACCTCTGCAAAATTCAGTGAAGTGAAGAAACCTGCGAGG 638  
QY 734 CAGTAACTCTGGGAGCTGGGAACCATGAAATTAAGACCCCAATATCANTCTGCGGAAG 793  
DB 639 CAGTAACTCTGGGAGCTGGGAACCATGAAATTAAGACCCCAATATCANTCTGCGGAAG 698  
QY 794 AATTGAGAAATATCTGTCCACCCCTCTCTCTCAGAGCCGAGAGTCCAGGCCCCCA 853  
DB 699 AATTGAGAAATATCTGTCCACCCCTCTCTCTCAGAGCCGAGAGTCCAGGCCCCCA 758  
QY 854 CCGCTCTCTCTTAAACCAAGAGGTACAGATCCAGCCCTCTCTCTTACAGACCGAGAG 913  
DB 759 CCGCTCTCTCTTAAACCAAGAGGTACAGATCCAGCCCTCTCTCTTACAGACCGAGAG 818  
QY 914 TCCAGACCCCGGCGCT 973

DB 819 TCCAGACCCCGGCGCT 878  
QY 974 CAGAGAGTCCACACCCCGGCGCT 1033  
DB 879 GAGAGGTCCACACCCCGGCGCT 938  
QY 1034 CCT 1093  
DB 939 CCT 998  
QY 1094 NNAGGTCCAGCCCT 1153  
DB 999 NNAGGTCCAGCCCT 1058  
QY 1154 NACAGATGCCCCCTGTGTGNGNAGTTGACCCCACTTACAGTGGTCTTTCATTTTGG 1213  
DB 1059 TACACAGTCCCCCTGTGTGNGNAGTTGACCCCACTTACAGTGGTCTTTCATTTTGG 1118  
QY 1214 TCCCTTCTCCCTAGATCCAGAAATTAAGTTTAAGANGCAAAAAA 1262  
DB 1119 TCCCTTCTCCCTAGATCCAGAAATTAAGTTTAAGANGCAAAAAA 1167

RESULT 7  
V11855  
ID V11855 standard; cDNA; 1386 BP.  
AC V11855:  
DE 11-SEP-1998 (first entry)  
KW Homo sapiens Tub Interactor (hTI-1) gene.  
KW serine protease; tub interactor; treatment; obesity; cachexia;  
KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;  
KW neurodegenerative disease; Alzheimer's disease; drug screening;  
KW Parkinson's disease; Huntington's chorea; detection; diagnosis;  
KW amyotrophic lateral sclerosis; spinocerebellar degeneration; ss.  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT 2..701  
FT CDS  
FT /tag= a  
FT /product= hTI-1 protein  
FT /note= "putative serine protease"  
PN W09812302-A1.  
PD 26-MAR-1998.  
PE 05-SEP-1997; U15627.  
PR 21-JUL-1997; US-897340.  
PR 17-SEP-1996; US-715032.  
PA (MILL-) MILLENNIUM PHARM INC.  
PI Errata PR; Glimeo CU;  
DR P-PsDB; M59129.  
PT Tub interactor genes - used to develop products for the treatment  
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.  
PT diabetes  
PS Claim 10; Fig 1; 120pp; English.  
CC The sequence is that of the Tub Interactor gene hTI-1 which  
CC codes for a putative serine protease. It genes function  
CC in biochemical pathways involved in weight control and  
CC related disorders. The products can be used for treating  
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,  
CC or a related disorder such as diabetes. The products can  
CC also be used to modulate cell cycle progression and apoptosis.  
CC They can be used for treating neurodegenerative diseases  
CC which are characterised by apoptosis, including Alzheimer's  
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic  
CC lateral sclerosis or spinocerebellar degenerations. The  
CC products can also be used for detection, diagnosis and  
CC drug screening.  
SQ Sequence 1386 BP; 318 A; 490 C; 321 G; 249 T;

Query Match 68.9%; Score 871.6; DB 1; Length 1386;  
Best Local Similarity 81.6%; Pred. No. 3.8e-203;  
Matches 1147; Conservative 3; Mismatches 25; Indels 230; Gaps 4;



```

OY 8 CGGACTCGAGCCCTGGCAGGCGGCGACTGATGAGAAAAGAAATGTTCTGCTGGGCG 67
    |||||||
DB 15 CGGACTCGAGCCCTGGCAGGCGGCGACTGATGAGAAAAGAAATGTTCTGCTGGGCG 74
OY 68 TCCTGTGATCGGCACTGGGCTGTCTCAGCCGACACTGTTTCCAGAA----- 117
    |||||||
DB 75 TCCTGTGATCGGCACTGGGCTGTCTCAGCCGACACTGTTTCCAGAAATGATGAGCA 134
OY 117 -CTCTACATCGGCTGGGCTGACAGTCTTGAGGCGACCAAGAGCGAGGAGCC 175
    |||||||
DB 135 GCTCTTACACATCGGCTGGGCTGACAGTCTTGAGGCGACCAAGAGCGAGGAGCC 194
OY 176 AGATGTGAGGCGAGGCTCTCCGTACGCGACCCAGAGATACAAAGACCTTGCTGCTA 235
    |||||||
DB 195 AGATGTGAGGCGAGGCTCTCCGTACGCGACCCAGAGATACAAAGACCTTGCTGCTA 254
OY 236 ACAGACTCATGCTCATATAGTTGGAGCAATCCGTTCCAGTCTGACACATCCGAGCA 295
    |||||||
DB 255 ACAGACTCATGCTCATATAGTTGGAGCAATCCGTTCCAGTCTGACACATCCGAGCA 314
OY 296 TCAGCATGCTTGGCAGTGGCTTACCGGCGGGAACCTTGCTGCTGCTGGGCTG 355
    |||||||
DB 315 TCAGCATGCTTGGCAGTGGCTTACCGGCGGGAACCTTGCTGCTGCTGGGCTG 374
OY 356 TGGTGGGGAACGGTGAAGCTCAGGGGTGTGTCTGCTTCAAGGAGGTCTCTGCC 415
    |||||||
DB 375 TGGTGGGGAAC----- 386
OY 416 AGTCCGCGGGGGCTGAGCCAGAGCTCTGCTCCAGCGAGCAATGCTTACGCTGCA 475
    |||||||
DB 386 -----GGAGATAGCTTACCTGCTGCAAGT 411
OY 476 CGTGAACGTGTCGGTGTGTCTGAGAGGTCTGAGTATAGTATACCGGCTGACCA 535
    |||||||
DB 412 GGTGAACGTGTCGGTGTGTCTGAGAGGTCTGAGTATAGTATACCGGCTGACCA 471
OY 536 CCCGAGCATGTTTGCCTGCGCGGAGGCGCAAGACAGAGACTCTGCAACGGTGA 595
    |||||||
DB 472 CCCGAGCATGTTTGCCTGCGCGGAGGCGCAAGACAGAGACTCTGCAACGGTGA 531
OY 596 TGGGGGGGCTGATCTGCAACGGGTACTTGGAGGCTTGTCTTGGAAAAAGCCC 655
    |||||||
DB 532 TGGGGGGGCTGATCTGCAACGGGTACTTGGAGGCTTGTCTTGGAAAAAGCCC 591
OY 656 GTGGGCGCAAGTTGGGCTGCTGAGTCTACACCACTCTGCAAAATTCAGTGTGAT 715
    |||||||
DB 592 GTGGGCGCAAGTTGGGCTGCTGAGTCTACACCACTCTGCAAAATTCAGTGTGAT 651
OY 716 AGAGAAAAAGCTT-CCAGGCGCAAGTTAAGTGGGAGTGGGAAACCATGAAATGACCCC 774
    |||||||
DB 652 AGAGAAAAAGCTTACAGGAGGAGTAACTGTGGGAGTGGGAAACCATGAAATGACCCC 711
OY 775 AATATACCTCTGCGGAGGAATTCAGAAATATCTGTTCCAGGCCCTCTCTCAGGCC 834
    |||||||
DB 712 AATATACCTCTGCGGAGGAATTCAGAAATATCTGTTCCAGGCCCTCTCTCAGGCC 771
OY 835 CAGGAGTCAAGGCGGCGGCTGCTGCTCAAAACCAAGGCTCAATCCCGAGCCCT 894
    |||||||
DB 772 CAGGAGTCAAGGCGGCGGCTGCTGCTCAAAACCAAGGCTCAATCCCGAGCCCT 831
OY 895 CTTCTCTCAGACCCAGAGTCTCAGACCCCGAGCCCTCTCTCAGACCCAGAGTCC 954
    |||||||
DB 832 CTTCTCTCAGACCCAGAGTCTCAGACCCCGAGCCCTCTCTCAGACCCAGAGTCC 891
OY 955 AGCCCTCTCTCAGACCCAGAGTCTCAGAC----- 988
    |||||||
DB 892 AGCCCTCTCTCTCTCAGACCCAGAGTCTCAGACCCCGAGCCCTCTCTCTCAGACCA 951
OY 988 ----- 988
DB 952 GGGGTCAAGCTCTCTCTCTCAGACCCAGAGTCTCAGACCCCGAGCCCTCTCTCTC 1011
OY 988 -----CCCCAGCCCTC 1000

```

```

DB 1012 AGACCCAGAGTCCAGCCCTCTCTCTCAGACCCAGAGATCCAGATCCAGCCCGAGCCCTC 1071
OY 1001 CTCCTCTCAGACCCAGAGGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1060
DB 1072 CTCCTCTCAGACCCAGAGGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1131
OY 1061 GCGCCCAACCCCTGTTTCCGACAGCCAGAGGTTNAGTCCAGCCGCGGCTCTCTCAGAGA 1120
DB 1132 GCGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1191
OY 1121 CCCAGAGTCCAGTCCAGTATGCTTCCCTGTCAGAGTCCGCGGCTCTCTCTCTCTCTCTCT 1180
DB 1192 CCCAGAGTCCAGTCCAGTATGCTTCCCTGTCAGAGTCCAGGCGGCGGCGGCGGCGGCGG 1251
OY 1181 ACCCAACTTACAGTGGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1240
DB 1252 ACCCAACTTACAGTGGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1311
OY 1241 GTTTAAGAGANGCAAAAAAAA 1265
DB 1312 GTTTAAGAGANGCAAAAAAAA 1336

RESULT 8
V37495
ID V37495 standard; DNA: 871 bp.
AC V37495;
DT 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
OS Homo sapiens.
FH key Location/Qualifiers
FT CDS
    31..777
    /tag= a
    /transl_except- (pos:367..369, aa:Xaa)
    /transl_except- (pos:412..414, aa:Xaa)
    /transl_except- (pos:424..426, aa:Xaa)
    /product= "HPSK protein"
    /note= "Xaa = unknown"
FT FT
    MO9820117-A1.
    PD 14-MAY-1998.
    PF 31-OCT-1997; U20051.
    PR 05-NOV-1996; US-744026.
    PA (INCY-) INCYTE PHARM INC.
    PI Bandman O, Goli SK;
    DR WPI: 98-286933/25.
    DR P-PSDB: M60592.
    PT New isolated prostate-specific kallikrein - used to develop products
    for diagnosis and treatment of, e.g. prostate carcinoma or benign
    hyperplasia.
    PS Claim 5; Fig 1A-C; 68bp; English.
    CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
    containing an expression vector comprising the HPSK nucleic acid sequence
    can be used to produce the protein recombinantly. The HPSK products can
    be used for the diagnosis of conditions or diseases associated with
    expression of HPSK such as prostate carcinoma and benign prostate
    hyperplasia. Agonists and antagonists which specifically bind to HPSK and
    modulate its activity can be used for the preparation of treatment of
    such conditions or diseases. The products can also be used for detection
    and drug screening, especially for the detection of prostate-specific
    kallikrein (PSK).
    SQ Sequence 871 bp; 166 A; 260 C; 258 G; 184 T;

Query Match 50.7%; Score 641; DB 1; Length 871;
Best Local Similarity 89.0%; Pred. No. 3 9e-147;
Matches 737; Conservative 0; Mismatches 8; Indels 83; Gaps 1;
DB 127 CAGCATTCAGCCCTGGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 186

```

QY 67 GTCTGTCATCCGAGTGGGTGTCTGTACGCCGACACTGTTTCCAGAACTCTCAAC 126  
 |||||  
 Db 187 GTCTGTCATCCGAGTGGGTGTCTGTACGCCGACACTGTTTCCAGAACTCTCAAC 246  
 QY 127 ATCGGCTGGGCTGTACAGTCTTGTAGGCGGACCAAGAGCCAGGAGCCAGATGTGGAG 186  
 |||||  
 Db 247 ATCGGCTGGGCTGTACAGTCTTGTAGGCGGACCAAGAGCCAGGAGCCAGATGTGGAG 306  
 QY 187 GCGAGCTCTCCCTAGGCGACAGAGTACAAAGACCTTCTCTCAACGACCTCAAG 246  
 |||||  
 Db 307 GCGAGCTCTCCCTAGGCGACAGAGTACAAAGACCTTCTCTCAACGACCTCAAG 366  
 QY 247 CTCATCAAGTTGGAGCAATCCGTGCGAGTGTACACCATCCGAGCATCACTATGCT 306  
 |||||  
 Db 367 CTCATCAAGTTGGAGCAATCCGTGCGAGTGTACACCATCCGAGCATCACTATGCT 426  
 QY 307 TCGCACTGCCCTACCGCGGGGAACTCTTGGCTGTTTGGCTGGGCTGTGTGGCGAAC 366  
 |||||  
 Db 427 TCGCACTGCCCTACCGCGGGGAACTTCTTGGCTGTTTGGCTGGGCTGTGTGGCGAAC 486  
 QY 367 GGTGACTCAACGGGTGTGTGTCTGCTCTTCAAGAGGCTCTCTGCCCCAGTGGGGGG 426  
 |||||  
 Db 487 ----- 487  
 QY 427 CTGACCCAGAGCTGTGCTGCCAGGAGAAATGCTACCGTGTGCTGACGTGACAGTGT 486  
 |||||  
 Db 487 -----GCGAGAAATGCTGTACCGTGTGCTGACGTGACAGTGT 523  
 QY 487 CCGTGTGTCTGAGAGGTCTGTGAGTATGATACCGGCTGTACACCCAGCATGT 546  
 |||||  
 Db 524 CCGTGTGTCTGAGAGGTCTGTGAGTATGATACCGGCTGTACACCCAGCATGT 583  
 QY 547 TCGGCGCGCGGAGGAGGCAAGACAGAGACTCTCTGCAACGGTGTGCTGGGGGGCCCC 606  
 |||||  
 Db 584 TCGGCGCGCGGAGGAGGCAAGACAGAGACTCTCTGCAACGGTGTGCTGGGGGGCCCC 643  
 QY 607 TGAATCTGAACAGGGTACTGTGACAGGGCTTGTCTTCCGAAAAGCCCGTGGGCGAAG 666  
 |||||  
 Db 644 TGAATCTGAACAGGGTACTGTGACAGGGCTTGTCTTCCGAAAAGCCCGTGGGCGAAG 703  
 QY 667 TTGGCGTGCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATAGAAAAACCG 726  
 |||||  
 Db 704 TTGGCGTGCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATAGAAAAACCG 763  
 QY 727 TCCAGGCGAGTTAACTCTGGGGAGTGGGAACCATGAATTAATGACCCCAATTCATCTG 786  
 |||||  
 Db 764 TCCAGGCGAGTTAACTCTGGGGAGTGGGAACCATGAATTAATGACCCCAATTCATCTG 823  
 QY 787 CCGAAGGAATTCAGAAATATCTGTTCCAGCCCTCTCCCTCAGGCC 834  
 |||||  
 Db 824 CCGAAGGAATTCAGAAATATCTGTTCCAGCCCTCTCCCTCAGGCC 871  
 RESULT 9  
 V58646  
 ID V58646 standard; cDNA: 1459 BP.  
 AC V58646;  
 DT 08-DEC-1998 (first entry)  
 DE Prostate tumour specific gene clone DB6.  
 KW Prostate tumour specific gene; human; prostate cancer; detection;  
 OS Homo sapiens.  
 PN M09837418-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998; U03690.  
 PR 09-FEB-1998; US-904809.  
 PR 25-FEB-1997; US-806596.  
 PR 01-AUG-1997; US-904809.  
 PA (CORI-) CORIXA CORP.  
 PI Dillion DC, Xu J.  
 DR WPI: 98-480805/41.  
 FT Novel human prostate specific tumour protein and fragments - useful

PR for detecting and treating prostate cancers  
 PS Claim 1: Page 114: 141pp: English.  
 CC This sequence represents a human prostate tumour specific gene, and can  
 CC be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprises contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC encoded by this sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 SQ Sequence: 1459 BP; 427 A; 328 C; 406 G; 295 T;  
 Query Match 37.5%; Score 474; DB 1: Length 1459;  
 Best Local Similarity 98.4%; Pred. No. 1.8e-106;  
 Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
 QY 104 ACTGTTCCAGAACTCTACACCATCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 163  
 |||||  
 Db 25 AGTGAAGTGAAGAGTCTCTACACCATCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAG 84  
 QY 164 AGCAGGAGCCAGATGTGTGAGAGCCAGCTCTCCGTAAGGCAACCAAGTACAAAGAC 223  
 |||||  
 Db 85 AGCAGGAGCCAGATGTGTGAGAGCCAGCTCTCCGTAAGGCAACCAAGTACAAAGAC 144  
 QY 224 CCGTGTGCTTACCAACCTCATGCTCATCAAGTGTGAGAGATCCGTGCCAGTGTACA 283  
 |||||  
 Db 145 CCGTGTGCTTACCAACCTCATGCTCATCAAGTGTGAGAGATCCGTGCCAGTGTACA 204  
 QY 284 CCATCCGAGCATACGATGCTTGCAGTGTGCTTGCAGCGGAGAACTTGTGCTGTT 343  
 |||||  
 Db 205 CCATCCGAGCATACGATGCTTGCAGTGTGCTTGCAGCGGAGAACTTGTGCTGTT 264  
 QY 344 CTGGCTGGGGTGTCTGTGGGAAAGGTGAGCTACAGGGTGTGTGCTGCTCTTCAAGA 403  
 |||||  
 Db 265 CTGGCTGGGGTGTCTGTGGGAAAGGTGAGCTACAGGGTGTGTGCTGCTCTTCAAGA 324  
 QY 404 GGTCTCTGCCCCAGTCCGCGGGGGGTGACCCAGAGCTCTGCTGCCAGCAATGCTTAC 463  
 |||||  
 Db 325 GGTCTCTGCCCCAGTCCGCGGGGGGTGACCCAGAGCTCTGCTGCCAGCAATGCTTAC 384  
 QY 464 CGTGTGCAAGTCTGTAACAGTGTGCGTGTGTGAGAGGCTCTCAATAGCTTATGA 523  
 |||||  
 Db 385 CGTGTGCAAGTCTGTAACAGTGTGCGTGTGTGAGAGGCTCTCAATAGCTTATGA 444  
 QY 524 CCGGCTGTACACCCCGAGATGTTCTGCGCGGCGGAGGCAAGACGAAGAGACTCTG 583  
 |||||  
 Db 445 CCGGCTGTACACCCCGAGATGTTCTGCGCGGCGGAGGCAAGACGAAGAGACTCTG 504  
 QY 584 CAACG 588  
 |||||  
 Db 505 CAACG 509  
 RESULT 10  
 V61251  
 ID V61251 standard; cDNA: 1459 BP.  
 AC V61251;  
 DT 06-JAN-1999 (first entry)  
 DE cDNA sequence of prostate tumour clone P703 splice variant DB6.  
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
 OS Homo sapiens.  
 PN M09837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998; U03492.  
 PR 09-FEB-1998; US-020956.  
 PR 25-FEB-1997; US-806099.  
 PR 01-AUG-1997; US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillion DC, Xu J.  
 DR WPI: 98-609886/51.  
 FT Polypeptides comprising immunogenic portions of prostate proteins -

PT used in a vaccine for the treatment of prostate cancer  
PS Claim 3; Page 106; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;  
  
Query Match 37.5%; Score 474; DB 1; Length 1459;  
Best Local Similarity 98.4%; Pred. No. 1.8e-106;  
Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;  
  
QY 104 ACGTTTCCAGAACTCTTACCACTCGGGCTGGGCTGCACAGTCTTGAGGCCACCAAG 163  
DB 25 AGTAGAGCAGAGCTCTTACCACTCGGGCTGGGCTGCACAGTCTTGAGGCCACCAAG 84  
QY 164 AGCCAGGAGCAGATGATGAGAGCCAGCCCTCTCCGACGACCCAGAGTACAACAGAC 223  
DB 85 AGCCAGGAGCAGATGATGAGAGCCAGCCCTCTCCGACGACCCAGAGTACAACAGAC 144  
QY 224 CTTGTCTGCTAACGACCTCATGCTCATCAAGTGAAGCAATCCGTCTCGAGTCTGACA 283  
DB 145 CTTGTCTGCTAACGACCTCATGCTCATCAAGTGAAGCAATCCGTCTCGAGTCTGACA 204  
QY 284 CCATCCGAGATAGCATGATGCTTCCGAGTCCCTACCGGGGAGAACTCTTCTCTGTT 343  
DB 205 CCATCCGAGATAGCATGATGCTTCCGAGTCCCTACCGGGGAGAACTCTTCTCTGTT 264  
QY 344 CTGAGCTGGGCTCTGCTGGGCAAGCGTACGAGTGTGTCTGTCTGCTTCAAGGA 403  
DB 265 CTGAGCTGGGCTCTGCTGGGCAAGCGTACGAGTGTGTCTGTCTGCTTCAAGGA 324  
QY 404 GGTCTCTGCCAGTCCGCGGGGGCTGAGCCAGAGCTGTGCTCCAGGACAGATGCTTAC 463  
DB 325 GGTCTCTGCCAGTCCGCGGGGGCTGAGCCAGAGCTGTGCTCCAGGACAGATGCTTAC 384  
QY 464 CGTCTCTGAGTGTGCTGAACTGTGCTGTCTGAGAGAGTCTGCAAGTCTATGA 523  
DB 385 CGTCTCTGAGTGTGCTGAACTGTGCTGTCTGAGAGAGTCTGCAAGTCTATGA 444  
QY 524 CCGGCTGACACCCAGCATGTCTGTGGCGGGGGAGAGGCAAGACAGAGACTCTG 583  
DB 445 CCGGCTGACACCCAGCATGTCTGTGGCGGGGGAGAGGCAAGACAGAGACTCTG 504  
QY 584 CAACG 588  
DB 505 CAACG 509  
  
RESULT 11  
V58648  
ID V58648 standard; cDNA; 1119 BP.  
AC V58648:  
DT 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE14.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
KM therapy; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 34..558  
FT FT /\*tag= a  
FT PN MO9837418-A2.  
FT PD 27-AUG-1998.  
FT PE 25-FEB-1998: US-904809.  
FT PR 09-FEB-1998: US-904809.  
FT PR 25-FEB-1997: US-806596.  
FT PR 01-AUG-1997: US-904809.  
FT PA (CORI-) CORIXA CORP.  
FT PI Dillon DC, Xu J;  
FT DR WPI; 98-480805/41.

DR P-PSDB; W69389.  
PT Novel human prostate specific tumour protein and fragments - useful  
PT for detecting and treating prostate cancers  
PS Claim 1; Page 116-117; 141pp; English.  
CC This sequence represents a human prostate tumour specific gene, and can  
CC be used in the method of the invention. The method is for detecting  
CC prostate cancer comprising contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC encoded by this sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancers. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.  
SQ Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;  
  
Query Match 28.6%; Score 361.4; DB 1; Length 1119;  
Best Local Similarity 99.7%; Pred. No. 4e-79;  
Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 8 CGCAGTGGAGCCCTGCGAGCGGCGCACTGTCATGAAACGAATTTCTGCTCGGGCG 67  
DB 2 CGCAGTGGAGCCCTGCGAGCGGCGCACTGTCATGAAACGAATTTCTGCTCGGGCG 61  
QY 68 TCCGTGTGATCCGACAGTGGTGTGTCACCGCACACTGTTCCAGAACTCTACACA 127  
DB 62 TCCGTGTGATCCGACAGTGGTGTGTCACCGCACACTGTTCCAGAACTCTACACA 121  
QY 128 TCGGGCTGGGCTCTGACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGAGG 187  
DB 122 TCGGGCTGGGCTCTGACAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGAGG 181  
QY 188 CCAGCTCTCCGTACGCGGACCCAGAGTACAACAGACCTTCTGCTAAGAGCTATGC 247  
DB 182 CCAGCTCTCCGTACGCGGACCCAGAGTACAACAGACCTTCTGCTAAGAGCTATGC 241  
QY 248 TCATCAAGTTGAGAGATCCGTGTCGAGTCTGACACATCCGAGATGAGATTCCTT 307  
DB 242 TCATCAAGTTGAGAGATCCGTGTCGAGTCTGACACATCCGAGATGAGATTCCTT 301  
QY 308 CGCAGTCCCTTACCGCGGGGAACTTTCCTGTTTCTGAGTGGGCTCTGCGGCAAGC 367  
DB 302 CGCAGTCCCTTACCGCGGGGAACTTTCCTGTTTCTGAGTGGGCTCTGCGGCAAGC 361  
QY 368 GTG 370  
DB 362 ATG 364  
  
RESULT 12  
V61253  
ID V61253 standard; cDNA; 1119 BP.  
AC V61253:  
DT 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DE14.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 34..558  
FT FT /\*tag= a  
FT PN MO9837093-A2.  
FT PD 27-AUG-1998.  
FT PE 25-FEB-1998: US-90492.  
FT PR 09-FEB-1998: US-020956.  
FT PR 25-FEB-1997: US-806099.  
FT PR 01-AUG-1997: US-904804.  
FT PA (CORI-) CORIXA CORP.  
FT PI Dillon DC, Xu J;  
FT DR WPI; 98-609886/51.  
FT DR P-PSDB: W71873.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PT used in a vaccine for the treatment of prostate cancer  
PS Claim 3; Page 108-109; 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA

CC was identified by analysis of a subextracted cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SO Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 28.6%; Score 361.4; DB 1; Length 1119;  
 Best Local Similarity 99.7%; Pred. No. 4e-79;  
 Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CCGACTCGCAGCCCTGGCAGCGCGGCTGTCATGAAAAGCAATTGTTCTGCTCGGGCG 67  
 DB 2 CCGACTCGCAGCCCTGGCAGCGCGGCTGTCATGAAAAGCAATTGTTCTGCTCGGGCG 61  
 QY 68 TCGTGGGATCGCGAGTGGGTGCTGTCAGCCGACACTGTTCCAGAACTCTTACACCA 127  
 DB 62 TCGTGGGATCGCGAGTGGGTGCTGTCAGCCGACACTGTTCCAGAACTCTTACACCA 121  
 QY 128 TCGGGCTGGGCTGACAGTCTTGAGGCCAGCAAGAGCCAGGAGCCAGATGTTGAGG 187  
 DB 122 TCGGGCTGGGCTGACAGTCTTGAGGCCAGCAAGAGCCAGGAGCCAGATGTTGAGG 181  
 QY 188 CCGAGCTCTCGTACGCGCAGCAGATGACAAAGACCTTGTGCTACGACCTCATGC 247  
 DB 182 CCGAGCTCTCGTACGCGCAGCAGATGACAAAGACCTTGTGCTACGACCTCATGC 241  
 QY 248 TCAATCAAGTTGAGCAATCCGTCGAGTCTGACACATCCGAGAGCATAGATGCTT 307  
 DB 242 TCAATCAAGTTGAGCAATCCGTCGAGTCTGACACATCCGAGAGCATAGATGCTT 301  
 QY 308 CCGAGTGGCTTACCGCGGGGAACTCTTGTGCTGCTGGGCTGCTGCTGCGGCAAGC 367  
 DB 302 CCGAGTGGCTTACCGCGGGGAACTCTTGTGCTGCTGGGCTGCTGCTGCGGCAAGC 361  
 QY 368 GTG 370  
 DB 362 ATG 364

## RESULT 13

X41114  
 ID X41114 standard; cDNA: 402 BP.

AC X41114;  
 DT 17-JUN-1999 (first entry)  
 DE Human secreted protein; EST: expressed sequence tag; diagnosis;  
 KW Human; secreted protein; EST: expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 OS Homo sapiens.  
 PN MO3906548-A2.  
 PD 11-FEB-1999.  
 PF 31-JUL-1998; IB1222.  
 PR 01-AUG-1997; US-905135.  
 PA (GSEST) GENSET.  
 PI Ductect A; Dumas M;line Edwards J, Lacroix B;  
 DR MPI: 99-153778/13.  
 DR P-PDB: Y12281.  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 PS Claim 1: Page 205; 824pp; English.  
 CC X41094 to X41347 represent 5 expressed sequence tags (ESTs) for human  
 CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
 CC respectively. The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductive hormone regulating activity, chemotactic/

CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell.  
 SO Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 27.4%; Score 346.8; DB 1; Length 402;  
 Best Local Similarity 96.4%; Pred. No. 1.1e-75;  
 Matches 370; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 2 GCAAGCCGCACTCGCAGCCCTGGCAGCGCGGCTGTCATGAAAAGCAATTGTTCTGCT 61  
 DB 17 GCAAGCCGCACTCGCAGCCCTGGCAGCGCGGCTGTCATGAAAAGCAATTGTTCTGCT 76  
 QY 62 CCGGGCTCGCTGGGATCGCGAGTGGGTGCTGTCAGCCGACACTGTTCCAGAA----- 117  
 DB 77 CCGGGCTCGCTGGGATCGCGAGTGGGTGCTGTCAGCCGACACTGTTCCAGAAAGTAG 136  
 QY 117 -----CTCTACACCATCGGGCTGCGCTGCTGACAGTCTTGAGGCCGCAAGAGCCAG 169  
 DB 137 TCAAGAGCTCTCAACACCATCGGGCTGCGCTGCTGACAGTCTTGAGGCCGCAAGAGCCAG 196  
 QY 170 GGAAGCAGATGTTGAGAGCCAGCTCTGCTACGCGCAGCAGATGACAAAGAGCTTGC 229  
 DB 197 GGAAGCAGATGTTGAGAGCCAGCTCTGCTACGCGCAGCAGATGACAAAGAGCTTGC 256  
 QY 230 TCGCTACACCATCGAGTCTGATCAAGTTGAGAGATCCGTCGAGTGTGACACATCC 289  
 DB 257 TCGCTACACCATCGAGTCTGATCAAGTTGAGAGATCCGTCGAGTGTGACACATCC 316  
 QY 290 GGAAGCAGATGTTGAGAGTCCCTACCGCGGGGAACTCTGCTGTTCTGCT 349  
 DB 317 GGAAGCAGATGTTGAGAGTCCCTACCGCGGGGAACTCTGCTGTTCTGCT 376  
 QY 350 GGGGTCTGCTGGCGAAGGTTGAGC 373  
 DB 377 GGGGTCTGCTGGCGAAGGTTGAGC 400

## RESULT 14

Q13332/C  
 ID Q13332 standard; DNA: 8174 BP.

AC Q13332;  
 DT 07-NOV-1991 (first entry)  
 DE GDP-Fuc:beta-D-galactoside alpha(1,2)-fucosyltransferase gene.  
 KW Glycosyltransferase.  
 OS Homo sapiens.  
 FH Key  
 FT cds Location/Qualifiers  
 FT /tag= a  
 PN MO9112340-A.  
 PD 22-AUG-1991.  
 PF 14-FEB-1991; U00899.  
 PR 14-FEB-1990; US-480133.  
 PR 14-FEB-1990; US-479858.  
 PR 12-DEC-1990; US-627621.  
 PA (UNMI) UNIV OF MICHIGAN.  
 PI Lowe JB: 1991; 155p; English.  
 DR MPI: 91-267151/36.  
 DR P-PDB: R13751.  
 PT Isolation of gene conveying post-translational characteristic -  
 PT e.g. the presence of soluble or membrane bound oligo or  
 PT polysaccharide or glycosyltransferase.  
 PS Disclosure: Fig 3; 155pp; English.  
 CC The DNA encodes a protein sequence capable of functioning as a  
 CC GDP-Fuc:beta-D-Gal alpha(1,2)-fucosyltransferase. The sequence  
 CC coded by nucleotides 4782 - 5780 represents the functional protein.  
 CC The enzyme produced by the DNA sequence can be used in enzymatic



---

Db 187 GTCCCTGGTGCATCC

127 ATCGGGGCTGGGGCTGCACAGTCTGTAGGGCGGACCAAGAGCCAGGGAGGCCAGATGGTGGAG 186  
 247 ATCGGGGCTGGGGCTGCACAGTCTGTAGGGCGGACCAAGAGCCAGGGAGGCCAGATGGTGGAG 306  
 187 GCCAGCTCTCCGTAACGGGACCCAGAGTAAACAACAACCTTGTCTGTACAGACCTCATG 246  
 307 GCCAGCTCTCCGTAACGGGACCCAGAGTAAACAACAACCTTGTCTGTACAGACCTCATG 366  
 247 CTCATCAAGTTGGAGCAATCCGTTGCCAGTCTGACACCATCCGAGCATCAGATTGCT 306  
 367 ATCATCAAGTTGGAGCAATCCGTTGCCAGTCTGACACCATCCGAGCATCAGATTGCT 426  
 307 TCCAGATGCCCTACCGGGGGGAGCACTCTGGCCCTTTTGGCTGGGGGTTCTGTGGCCAAAC 366  
 427 TCCAGATGCCCTACCGGGGGGAGCACTTTGGCTCTGTTTCTGGGTTGGGTTCTGTGGCCAAAC 486  
 367 GGTGAGCTCACGGGTGTGTGTCTGCCCTCTTCAAGAGAGTCTCTGCCCCAGTCGGGGGGG 426  
 487 ----- 487  
 427 CTGACCCAGAGCTGTGCGTCCAGGACAGATGCTACCGTGTGTCAGTGGCTGAACGTGT 486  
 487 -----GGCAGAAATGCTACACGTCGTGTCAGTGGCTGAACGTGT 523  
 487 CCGTGTGTCAGTGGAGAGTCTGCACTAAGCTCTATAGACCCGCTGTACCAACCCACAGATGT 546  
 524 CCGTGTGTCAGTGGAGAGTCTGCACTAAGCTCTATAGACCCGCTGTACCAACCCACAGATGT 583  
 547 TCTGCCGCCGCGGAGGAGGACAGACAGAAAGAGTCTCTGCAACGGTACTCTGGGGGGGGCCC 606  
 584 TCTGCCGCCGCGGAGGAGGACAGACAGAAAGAGTCTCTGCAACGGTACTCTGGGGGGGGCCC 643  
 607 TGATTCGACAGCGGTACTTTCAGGGGCTTGTGTCTTTGGAAAAAGCCCCGTTGTGGCCAAAG 666  
 644 TGATTCGACAGCGGTACTTTCAGGGGCTTGTGTCTTTGGAAAAAGCCCCGTTGTGGCCAAAG 703  
 667 TTGGGCTGTCAGAGTCTCTACACCAACCTCTGCAAAATTCATCTAGATGATAGAAAAACCG 726  
 704 TTGGGCTGTCAGAGTCTCTACACCAACCTCTGCAAAATTCATCTAGATGATAGAAAAACCG 763  
 727 TCCAGGCGCAGTAACTCTGGGGGACTGGGAACCCATGAATTAAGCCCCCAAAATACATCTGT 786  
 764 TCCAGGCGCAGTAACTCTGGGGGACTGGGAACCCATGAATTAAGCCCCCAAAATACATCTGT 823  
 787 CGGAAGGAATTCAGGAATATCTGTTCCAGGCCCTCTCTCCCTCAGGCC 834  
 834 CGGAAGGAATTCAGGAATATCTGTTCCAGGCCCTCTCTCCCTCAGGCC 871

RESULT 2  
 US-07-914-281-5/c  
 ; Sequence 5, Application US/07914281  
 ; Patent No. 5324663  
 ; GENERAL INFORMATION:  
 APPLICANT: LOWE, JOHN B.  
 TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
 TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
 TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
 TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURE  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: P.B.I., SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P.C.  
 STREET: 1755 Jefferson Davis Highway, Fourth Floor  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentln Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/914,281
FILING DATE: 19920720
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Lavalley, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-07-914-281-5

Query Match 14.6% Score 184.6; DB: 1; Length 8174;
Best Local Similarity 78.2%; Pred. No. 9.2e-38;
Matches 272; Conservative 0; Mismatches 64; Indels 12; Gaps 4;

QY 795 ATTCAGGATATCTTCTCCAGAGCCCTCTCCCTCAGGCGCCAGAGATCCAGGCCAGC 854
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 1025 ACTTAGAGATCCAGGCTCCGGCCCTCTCTCTAGACCCAGAGATCCAGGCCCTC 966
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 855 CCTCTCTCTCAACCAAGG-GTACAGATCCCGCCCTCTCTCTCAGACCCAGAG 913
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 965 CCTCTCTCTCAGACCCAGAGATCCAGAGCCCGCCCTCTCTCTCAGACCCAGAG 906
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 914 TCCAGACCCCCAGGCCCTCTCTCTCTAGACCCAGAGATCCAG-----CCCTCTC 965
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 905 TCCAGAA-TCCCTAGCCCTCTCTCTCTCTCAGACCCAGAGATCCAGACCAAGCTCCCTC 847
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 966 CNTCAGACCCAGAGATCCAGACCCCGCCCTCTCTCTCAGACCCAGAGGTTGAGGC 1025
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 846 CCTCAGACCCAGAGACCCAGAGTCCCGCCCTCTCTCTCAGATCCAGAGATCAGG- 787
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1026 CCCCAACCCCTCTCTCTCAGATCAGAGTCCAGAGCCCGCCCAACCCCTGTTCCCGAGAC 1085
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 787 -CCCGACCCCTCTCTCTCAGACCCAGAGATCCAGAGGCCCGCCCAACCCCTCTCTCAGAC 729
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
QY 1086 CCAGAGGTNNAGTCCCGAGCCCTCTCTCTCTCAGACCCAGAGTCCAA 1133
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||
DB 728 CCAGAGTCCAGAGCCCGCCCTCTCTCTCAGACACAGAGAGGCTTA 681
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||

RESULT 3
US-08-393-246-5/C
; Sequence No. 5595900
; Patent No. 5595900
; GENERAL INFORMATION:
; APPLICANT: LOWE, JOHN B.
; TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
; OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
; TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
; TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 Jefferson Davis Highway, Fourth floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```



```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,246
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-08-393-246-5

```

```

Query March          14.6% Score 184.6; DB 1; Length 8174;
Best Local Similarity 78.2%; Pred. No. 9,2e-38;
Matches 272; Conservative 0; Mismatches 64; Indels 12; Gaps 4;

OY 795 ATTGAGAAATATCTGTTCACAGCCCTCTCCTCAGGCCAGGAGTCCAGCCCCAGC 854
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1025 ACTTAGAGATCCAGGCTCTCGGGCCCTCTCTTCAGACCAGGAGTCCAGCCCCAGC 966
OY 855 CCGCTCCCTCCCTCAAACCAAGG-GTACAGATCCCAGCCCCCTCCTCCAGACCAGGAG 913
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 965 CCTCTCTCTCTCAGACCCAGGAGTCCAGACCCAGCCCCCTCTCTCTCACACCCAGAG 906
OY 914 TCCAGACCCCCAGCCCCCTCTCCTCTCAGACCAGGAGTCCAG-----CCCTCTCTC 965
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 905 TCCAGA-TCCCTAGCCCTACTCTCCTCAGACCAGGAGTCCAGACCAAAGCTCCTCTC 847
OY 966 CNTAGACCCAGGAGATCCAGACCCCCCAGCCCCCTCTCCTCTCAGACCAGGAGTTAGGC 1025
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 846 CCTCAGACCCAGGAGCCCAAGTTCCCGCCCTCTCTCTCAGATCCAGGAGTTCAGG- 787
OY 1026 CCCCAACCCCTCTCTCTCTCAGAGTCCAGAGTCCAAGCCCCCAACCCCTCTCTCCAGAC 1085
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 -CCGAGACCTCTCTCTCTCTCAGACCCAGGAGTCCAGAGCCCCCAGCCCTCTCTCCCTCAGAC 729
OY 1086 CCAGAGGTTNNAAGTGTCACAGCCCTCTTCCNTCAGACCCAGGAGTCCAA 1133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 CCAGAGTCCAGAGCCCGAGCCCTCTCTCTCAGACACAGAGAGGCTA 681

RESULT 4
US-08-525-058A-5/c
Sequence 5, Application US/08525058A
Patent No. 5770420
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLOK, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
```

```

STREET: 1755 Jefferson Davis Highway, Fourth Floor
City: Arlington
STATE: Virginia
COUNTRY: U.S.A.
Zip: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,058A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-08-525-058A.5

```

```

Query Match 14.6%; Score 184.6; DB 3; Length 8174;
Best Local Similarity 78.2%; Pred: No. 9.2e-38;
Matches 272; Conservative 0; Mismatches 64; Indels 12; Gaps 4;

QY 795 ATTCAGGAAATATCTGTTCCTCCAGGCCCTCTCTCCCTCAGGCCAGAGTCCAGGCCCCAGC 854
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1025 ACTTAGAGGTCCAGAGCTCCGGCCCTCTCTCTCTCAGACCAGAGTCCAGGCCCTCTCG 966

QY 855 CCGTCCTCTCCCTCAAAACCAAGG-GTACAGATCCCAAGCCCTCTCTCTCCTCAGACCAGAG 913
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 965 CCGTCCTCTCTCAGACCAGAGGTCCAGACCAGCCAGCCCTCTCTCTCCTACACCCAGAG 906

QY 914 TCACAGACCCCCAGCCCTCTCTCTCTAGACCAGAGAGTCCAG-----CCCTCTCTC 965
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 905 TCCAGA-TCCCTAACCCCTCTACTCTCTCAGACCAGAGTCCAGACCAGAGCTCTCTC 847

QY 966 CNTAGAGACCAGAGTCCAGACCACCCCGCCCTCTCTCTCTCAGACCAGAGGTTCAGGC 1025
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 846 CCTCAGACCCAGAGGCCCAAGTCCCGAGCCCTCTCTCTCTCAGATCCAGAGTACAGG- 787

QY 1026 CCCCACCCCTCTCTCTCAGAGTCCAGAGTCCCAAGCCCAAGCCCTCTCTCTCCAGAG 1085
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 -CCAGAGACCTCTCTCTCTCAGACCAGAGTCCAGAGCCCAAGCCCTCTCTCTCTAGAG 729

QY 1086 CCAGAGGTTNAGTCCAGGCCCTCTCTCTCTCAGAGACCAGAGGTGCCAA 1133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 CCAGAGGTCCAGAGCCCAAGCCCTCTCTCTCTCAGACACAGAGGCTTA 681

RESULT 5
PCT-US91-00899-3/C
: Sequence 3, Application PC/TUS9100899
:
: GENERAL INFORMATION:
:   APPLICANT: Lowe, John B.
:   TITLE OF INVENTION: Method and Products For the Synthesis of
:   TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,
:   TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned
:   TITLE OF INVENTION: Genetic Sequences That Determine These Structures
:   NUMBER OF SEQUENCES: 16
:   CORRESPONDENCE ADDRESS:
:

```

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/00899  
FILING DATE: 19910214  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Lavalleye Ph.D., Jean-Paul  
REGISTRATION NUMBER: 31,451  
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)521-5940  
TELEFAX: (703)486-2347  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8174 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
TISSUE TYPE: Blood  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 4686..5780  
OTHER INFORMATION: /label= mat\_peptide  
PCT-US91-00899-3

Query Match 14.6%; Score 184.6; DB 5; Length 8174;  
Best Local Similarity 78.2%; Pred. No. 9.2e-38;  
Matches 272; Conservative 0; Mismatches 64; Indels 12; Gaps 4;  
QY 795 ATTCAGGAATATCTGTCCAGGCGCTCCCTCAGGCGCCAGAGTCCAGGCGCCAGC 854  
DB 1025 ACTTAGAGTCCAGGCTCCGCGCTCTCTCTCAGACCCAGAGTCCAGGCGCCAGC 966  
QY 855 CCCTCTCTCTCAAAACCAAGG-GTACAGATCCCAAGCCCTCTCTCTCAGACCCAGAG 913  
DB 965 CCCTCTCTCTCAGACCCAGAGTCCAGACCCAGCCCTCTCTCTCAGACCCAGAG 906  
QY 914 TCCAGACCCCGCCCTCTCTCTCTCAGACCCAGAGTCCAG-----CCCTCTCTC 965  
DB 905 TCCAGA-TCCCTGTGCGCTCTCTCTCTCAGACCCAGAGTCCAGACCAAGACTCCTCTC 847  
QY 966 CNTCAGACCCAGAGTCCAGACCCAGGCGCTCTCTCTCTCAGACCCAGAGGTTAGGC 1025  
DB 846 CCTCAGACCCAGAGGCGCCAGAGTCCAGGCGCTCTCTCTCTCAGAGTCCAGAG- 787  
QY 1026 CCCCAGACCCCTCTCTCTCTCAGAGTCCAGAGTCCAGACCCAGCCCTCTCTCTCAGAG 1085  
DB 787 -CCCAAGCCCTCTCTCTCTCAGACCCAGAGTCCAGGCGCCCGCCAGCCCTCTCTCTCAGAG 729  
QY 1086 CCAAGAGTNNAGTCCAGGCGCTCTCTCTCTCAGACCCAGAGGTTCCAA 1133  
DB 728 CCAAGAGTCCAGAGCCCGCCAGGCGCTCTCTCTCTCAGACCAAGAGGCTTA 681

RESULT 6  
US-08-308-949A-1/c  
; Sequence 1; Application US/08308949A

Patent No. 5580703  
GENERAL INFORMATION:  
APPLICANT: Kotlin, Robert M.  
APPLICANT: Berns, Kenneth I.  
APPLICANT: Linden, Ralph M.  
TITLE OF INVENTION: Human Adeno-Associated Virus Integration  
TITLE OF INVENTION: Site DNA and Uses Thereof  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/308,949A  
FILING DATE: September 20, 1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/947,127  
FILING DATE: September 27, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Cartoll, Alice O.  
REGISTRATION NUMBER: 33,542  
REFERENCE/DOCKET NUMBER: ACC92-10F  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-861-6240  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4060 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-308-949A-1

Query Match 11.5%; Score 145.2; DB 1; Length 4060;  
Best Local Similarity 72.5%; Pred. No. 6.2e-28;  
Matches 243; Conservative 0; Mismatches 78; Indels 14; Gaps 4;  
QY 813 CCAGCCCTCTCTCTCTCAGGCGCCAGAGTCCAGGCGCCAGGCGCCCTCTCTCAACCA 872  
DB 4000 CCAACCCCTCTCTCTCTCAACCCAGAGGCGCCAGGCGCCCTCTCTCTCAATCAA 3941  
QY 873 AGGTAAGATCCCAAGCCCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCAGCCCT 932  
DB 3940 GGAATCAGAGGCGCCAGGCGCTCTCTCTCTCAGACCCAGAGGCGCGAG---GCCAGCCCG 3884  
QY 933 CTTCTCTCAGACCCAGAGTCCAG-----CCCTCTCTCTCTCAGACCCAGAGTCCA 984  
DB 3883 CTTCTCTCAGACCCAGAGTCCAGGCGCCAGGCGCCAGGCGCCCTCTCTCTCAGACCA 3824  
QY 985 GACCCCGAGCCCTCTCTCTCTCAGACCCAGAGGTTGAGGCGCCAGCCCTCTCTCTTC 1044  
DB 3823 G--GCCCGAGCCCTCTCTCTCTCAGACCCAGAGTCCAGGCGCCAGGCGCCCTCTCTC 3766  
QY 1045 AGAGTCAAGTCCAGGCGCCAGGCGCCCTCTCTCTCTCAGACCCAGAGTNNAGTCCAG 1104  
DB 3765 AGACCCAGAGGTCG-AGGCGCCAGGCGCCCTCTCTCTCTCAGACCCAGAGTCCAGGCGCCAG 3707  
QY 1105 CCCCTCTCTCTCAGACCCAGAGGTTCCATGCGAC 1139  
DB 3706 CCCCTCTCTCTCAAAACCCAGAGGCGCCAGGCGCC 3672



APPLICANT: Little, Sheila P.  
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND  
RELATED NUCLEIC ACIDS  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Eli Lilly and Company  
STREET: Lilly Corporate Center  
CITY: Indianapolis  
STATE: Indiana  
COUNTRY: United States of America  
ZIP: 46285  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US96/04294  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/416,257  
FILING DATE: 04-APR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Blalock, Donna K.  
REGISTRATION NUMBER: 38,082  
REFERENCE/DOCKET NUMBER: X9239  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-277-1090  
TELEFAX: 317-276-3861  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1089 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: mRNA  
PCT-US96-04294-3

Query Match 7.6%; Score 96.6; DB 5; Length 1089;  
Best Local Similarity 59.8%; Pred. No. 7.3e-16;  
Matches 162; Conservative 0; Mismatches 109; Indels 0; Gaps 0;  
QY 457 TGCCACCGGCTGACAGTGTGCGGTGCTGAGAGGTGCAAGTAAAGC 516  
DB 498 TTCCCTGACCTGATGAGTGTGCTGATGCTCAAGTCTTCCCCCGAGCTGACGAGG 439  
QY 517 TGTATGACCCGCTGTACACCCAGCATGTTCTGCGCCGCGGAGGCAAGACGAGG 576  
DB 438 TTACAGAGCACTACTGAAAAATTCATGTGCTGCGCTGGCATCCCCGACTCCAGAAAA 379  
QY 577 ACTCCGCAAGGTGACTGTGGGGGCCCCCTGATCTGCAACGGGTCTTGCAAGGCTTG 636  
DB 378 AGCCCTGCAATGTGTACTAGGGGGACCGTTGGTGTGCAAGGTACCTGCAAGGTCTG 319  
QY 637 TGTCTTGGAAAAAGCCGCTGTGGCCAAAGTTGGCGCAGGTGCTACCAACCTCT 696  
DB 318 TGTCTGGGGAAGTTCCTTGTGGGCAACCCCAATGACCAAGAGTCTCACTCAAGTCT 259  
QY 697 GCAATTCACGTAGATAGAGAAAAACGCT 727  
DB 258 GCAAGTTCACCAAGTGAATAATGACACCAT 228  
RESULT 10  
US-08-557-146-1  
Sequence 1, Application US/08557146  
Patent No. 5834290  
GENERAL INFORMATION:  
APPLICANT: Egelund, Torbjorn  
APPLICANT: Hansson, Lennart  
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic

TITLE OF INVENTION: Enzyme (SCCE)  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: White & Case, Patent Department  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2787  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/557,146  
FILING DATE: 14-DEC-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Sterner, Richard J.  
REGISTRATION NUMBER: 35,372  
REFERENCE/DOCKET NUMBER: 1103326-181  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 819-8783  
TELEFAX: (212) 354-8113  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 986 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 25..786  
FEATURE:  
NAME/KEY: s19\_peptide  
LOCATION: 25..90  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 112..783  
US-08-557-146-1

Query Match 7.5%; Score 95; DB 4; Length 986;  
Best Local Similarity 59.4%; Pred. No. 1.8e-15;  
Matches 161; Conservative 0; Mismatches 110; Indels 0; Gaps 0;  
QY 457 TGCCACCGTGTGCAAGTGTGCGGTGCTGAGAGGTGCAAGTAAAGC 516  
DB 500 TTCCCTGACCTGATGAGTGTGCTGATGCTCAAGTCTTCCCCCGAGCTGACGAGG 559  
QY 517 TGTATGACCCGCTGTACACCCAGCATGTTCTGCGCGGAGGCAAGACGAGG 576  
DB 560 TTACAGAGCACTTACTGAAAAATTCATGTGCTGCGCTGGCATCCCCGACTCCAGAAAA 619  
QY 577 ACTCCGCAAGGTGACTGTGGGGGCCCCCTGATCTGCAACGGGTACTGCAAGGCTTG 636  
DB 620 AGCCCTGCAATGTGTACTAGGGGGACCGTTGGTGTGCAAGGTACCTGCAAGGTCTG 679  
QY 637 TGTCTTGGAAAAAGCCGCTGTGGCCAAAGTTGGCGTCCAGAGTGTCTACACCAACCTCT 696  
DB 680 TGTCTGGGGAAGTTCCTTGTGGGCAACCCCAATGACCAAGAGTCTCACTCAAGTCT 739  
QY 697 GCAATTCACGTAGATAGAGAAAAACGCT 727  
DB 740 GCAAGTTCACCAAGTGAATAATGACACCAT 770

## RESULT 11

US-08-790-137-2

Sequence 2, Application US/08790137

Patent No. 5840871

GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.

APPLICANT: Goll, Surya K.

TITLE OF INVENTION: A NOVEL PROSTATE-ASSOCIATED

TITLE OF INVENTION: KALLIKREIN

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Drive

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/790,137

FILING DATE: Filed Herewith

CLASSIFICATION: 424

Prior Application Number:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0195 US

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 833 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-790-137-2

Query Match 6.9%; Score 86.8; DB 4; Length 833;

Best Local Similarity 53.9%; Pred. No. 2e-13;

Matches 178; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 410 CTGCCAGTGGCGGGGCTACCCAGAGCTCTGTGCTCCAGGCGATGCTTACCGTCT 469  
DB 458 CTGCCATTTTCGGGGGCGGAGACAGCTCCAGCCCACTTACGCTGCTCAGACCTT 517  
QY 470 GCAAGCGGTAGAGTGTGCTGTGTGAGAGGCTTCGAGTATGATGACCGCT 529  
DB 518 GCGATGGCCCAATCATCATGAGCAGCAGAGTGTGAGAGCCCTTACCCGGCAA 577  
QY 530 GTAACCAACCCAGCATGTTTGGCGGGGCGAGGCGAAGACCAAGAGACTCTGCAACG 589  
DB 578 CATCACAGACACCATGTGTGTGTCAGAGCTGACAGAGGGGCGCAAGACTCTGCAAGG 637  
QY 590 TGACTGTGGGGGGCCCTGATCTGCAAGGGTACTTCAAGGCTTGTCTTTCGAAA 649  
DB 638 TGACTCGGGGGGCTGTGTGTGTACAGTCTCTTCAAGGCTTGTCTTTCGAGGCA 697  
QY 650 AGCCCGCTGGCCCAAGTGTGCGGCGAGGCTGTACACCAACCTTGTCAATTCACCTGA 709  
DB 698 GGAATCCTGTGTGCAATCAACCGAAGCCTGTGTGTACACCAAGTGTGCAATTCATGTGGA 757  
QY 710 GTGGATAGAGAAAACCGTCCAGGCGCACTTA 739  
DB 758 CTGGATCCAGAGACGATGAAGAACTTA 787

## RESULT 12

US-08-738-413B-8

Sequence 8, Application US/08738413B

Patent No. 5821106

GENERAL INFORMATION:

APPLICANT: CHUNG, Kwang-Hoe

APPLICANT: KOH, You-Seok

APPLICANT: Hwang, Jae-Hoon

APPLICANT: Kim, Doo-Sik

APPLICANT: YUN, Yung-Dae

APPLICANT: MOON, Hong-Mo

TITLE OF INVENTION: A NOVEL CDNA OF DIRECT-ACTING

TITLE OF INVENTION: FIBRINOLYTIC SERINE PROTEASE

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby &amp; Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: US

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/738,413B

FILING DATE: October 23, 1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ludwig, S. Peter

REGISTRATION NUMBER: 25,351

REFERENCE/DOCKET NUMBER: 0136/0C539

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237

TELEX: 236687

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 699 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: Agkistrodon halys brevicaudus

IMMEDIATE SOURCE:

CLONE: Protease

US-08-738-413B-8

Query Match 5.8%; Score 73.8; DB 3; Length 699;

Best Local Similarity 63.0%; Pred. No. 3.5e-10;

Matches 114; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 540 AGCATGTTCTGCGCGGCGGAGGCGAAGACCAAGAGACTCTGCAACGCTGACTGGG 599  
DB 481 AGAACAATGTGTGTCAGAGTATCTGGAAGAGGCAAGATTCATGTAACGTGACTGGG 540  
QY 600 GGGCCCCGATCTGCAACGGGTAATTCGAGGGGCTTGTGTTCGAAAAAGCCCGTGT 659  
DB 541 GGACCCCTCATCTGTATGCAATTCAGAGGCGATTCATATGAGGGGCGCGTACTTGT 600  
QY 660 GGCCAAAGTGGCGGAGGTGTCTACCAACCTCTGCAAAATTCAGTGTGATAGAG 719  
DB 601 GCCCAACCGGTGAGCTGGCTCTTACCAAGGCTTTGATTAATATGATGATCA 660  
QY 720 A 720  
DB 661 A 661





---



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 11:33:07 ; Search time 161.06 seconds  
(without alignments)  
720.047 Million cell updates/sec

Title: US-09-030-606-173

Perfect score: 1265

Sequence: 1 GCGAGCCCGCCACTCGCAGCC.....AGANGNCMAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/lna/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/lna/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/lna/5C\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/lna/5D\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/lna/PCTUS9\_COMB.seq:\*  
6: /cgn2\_6/ptodata/2/lna/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64.1	50.7	871	3	US-08-744-026-2
2	184.6	14.6	8174	1	US-07-914-281-5
3	184.6	14.6	8174	1	US-08-393-246-5
4	184.6	14.6	8174	1	US-08-525-058A-5
5	184.6	14.6	8174	5	PCT-US91-00899-3
6	145.2	11.5	4060	1	US-08-308-949A-1
7	106.2	8.4	732	3	US-08-361-395-2
8	96.6	7.6	1089	5	PCT-US96-04294-1
9	96.6	7.6	1089	5	PCT-US96-04294-3
10	95	7.5	986	4	US-08-557-146-1
11	86.8	6.9	833	4	US-08-790-137-2
12	73.8	5.8	699	3	US-08-738-413B-8
13	68.8	5.4	957	3	US-08-684-862-11
14	66	5.2	7218	2	US-08-232-463-14
15	65.4	5.2	1454	3	US-08-467-155A-2
16	65.4	5.2	1454	4	US-08-628-198-2
17	65.4	5.2	1454	5	PCT-US96-07343-2
18	64.4	5.1	988	3	US-08-684-862-10
19	63.6	5.0	832	5	PCT-US95-06157-5
20	63.6	5.0	760	5	PCT-US95-06157-7
21	63.6	5.0	766	5	PCT-US95-06157-9
22	63.2	5.0	992	2	US-08-358-782D-13
23	63.2	5.0	1462	2	US-08-358-782D-14
24	63.2	5.0	1729	4	US-08-844-024-1
25	61.8	4.9	1096	3	US-08-684-862-8
26	61	4.8	1333	3	US-08-684-862-9
27	60.6	4.8	840	3	US-08-684-862-12
28	54.8	4.3	8174	1	US-07-914-281-5
29	54.8	4.3	8174	1	US-08-393-246-5
30	54.8	4.3	8174	3	US-08-525-058A-5
31	54.8	4.3	200	3	US-08-931-981A-1
32	54.8	4.3	8174	5	PCT-US91-00899-3
33	52	4.1	2259	4	US-08-845-998-3
34	50.4	4.0	2218	4	US-08-845-998-5
35	49.2	3.9	734	3	US-08-650-129-1
36	49.2	3.9	821	3	US-08-650-129-2
37	49.2	3.9	866	3	US-08-650-129-3

## ALIGNMENTS

38	47.4	3.7	12001	3	US-08-458-568A-11	Sequence 11, Appl
39	46	3.6	329	1	US-08-148-910-2	Sequence 2, Appl
40	46	3.6	970	1	US-08-148-910-3	Sequence 3, Appl
41	46	3.6	329	1	US-08-148-910-13	Sequence 13, Appl
42	46	3.6	2033	1	US-08-148-910-14	Sequence 14, Appl
43	46	3.6	329	2	US-08-448-937A-2	Sequence 2, Appl
44	46	3.6	970	2	US-08-448-937A-3	Sequence 3, Appl
45	46	3.6	329	2	US-08-448-937A-13	Sequence 13, Appl

RESULT 1  
US-08-744-026-2  
; Sequence 2, Application US/08744026  
; Patent No. 5786148

GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
TITLE OF INVENTION: A NOVEL PROSTATE-SPECIFIC  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESS: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: US  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/744,026  
FILING DATE: Herewith  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: Pf-0154 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ. ID NO. 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 871 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE: Consensus

Query Match 50.7%; Score 641; DB 3; Length 871;  
Best Local Similarity 89.0%; Pred. No. 5.6e-153;  
Matches 737; Conservative 0; Mismatches 8; Indels 83; Gaps 1;

QY	7	CCGACATCGAGCCCTGGCAGCGGACATGTCATGGAACGAATGTCCTCGGGC	66
DB	127	CACGATTCGAGCCGCGGCGGCGGACATGTCATGGAACGAATGTCCTCGGGC	186
QY	67	GTCCTGTCATCGCAGTCGTCGTCAGCCGACACATGTTCCAGAACTCCTACACC	126
DB	187	GTCCTGTCATCGCAGTCGTCGTCAGCCGACACATGTTCCAGAACTCCTACACC	246

127 ATCGGCGTGGGCTTCACAGCTTGTAGGGCGGCAAGAGCCAGGGAGGACAGTGGTGGAG 186  
 247 ATCGGCGTGGGCTTCACAGCTTGTAGGGCGGCAAGAGCCAGGGAGGACAGTGGTGGAG 306  
 187 GCCAGCCTCTCCGTCAGGACACCCAGAGTACACACAGACCCCTTGCCTACAGACCTCATG 246  
 307 GCCAGCCTCTCCGTCAGGAGACCCAGAGTACACACAGACCCCTTGCCTACAGACCTCATG 366  
 247 CTCATCAAGATTGGAGCAATCCGTGTCCGAGTGTACACACCATCCGAGCATCAGCATTTCT 306  
 367 ATCATCAAGATTGGAGCAATCCGTGTCCGAGTGTACACACCATCCGAGCATCAGCATTTNT 426  
 307 TCGCAGTGGCCCTACCGGGGGGAACTTGTGCCCTGTTCCGGCTGGGGGTCTGTGGCGAAC 366  
 427 TCGCAGTGGCCCTACCGGGGGGAACTTGTGCCCTGTTCCGGCTGGGGGTCTGTGGCGAAC 486  
 367 GGTGAGCTCAGCGGTGTGTGTCTGCCCTTTTCAGAGAGGTCTCTGCCAGTCCGGGGGG 426  
 487 ----- 487  
 427 CTGACCCAGAGCTTGCGTCCAGGACAGAGTCTTACCGTGTCTGATGCGTGAACGTGT 486  
 487 -----GGCAGAAATGCTTACCGTGTCTGATGCGTGAACGTGT 523  
 487 -----GGCAGAAATGCTTACCGTGTCTGATGCGTGAACGTGT 523  
 487 CGGTGTGTCTGAGAGGCTGCACTAAGCTCTATAGACCCGTGTACACCCACACATGT 546  
 524 CGGTGTGTCTGAGAGGCTGCACTAAGCTCTATAGACCCGTGTACACCCACACATGT 583  
 547 TCTGCGCCGCGGCGAGGAGACACAGAAAGCATCTCTGCAACCGTGTACTGTGGGGGGCCCC 606  
 584 TCTGCGCCGCGGCGAGGAGACACAGAAAGCATCTCTGCAACCGTGTACTGTGGGGGGCCCC 643  
 607 TGATCTGCAACGGGTACTTTCAGGGGCTTGTGTCTTTTGGAAAAAGCCCCGTGTGGCCAA 666  
 644 TGATCTGCAACGGGTACTTTCAGGGGCTTGTGTCTTTTGGAAAAAGCCCCGTGTGGCCAA 703  
 667 TTGGCGTCCGAGAGTGTCTACACCAACCTCTGTGAATTCTACATGATGATAGAAAAACCG 726  
 704 TTGGCGTCCGAGAGTGTCTACACCAACCTCTGTGAATTCTACATGATGATAGAAAAACCG 763  
 727 TCCAGGCGCATTAATCTGAGGAGCTGGGAGCCATGAATTAAGCCCAATATACATCTGT 786  
 764 TCCAGGCGCATTAATCTGAGGAGCTGGGAGCCATGAATTAAGCCCAATATACATCTGT 823  
 787 CGGAAGGAATTCAGAAATATCTGTTCCAGGCCCTCTCTCCCTCAGGCC 834  
 824 CGGAAGGAATTCAGAAATATCTGTTCCAGGCCCTCTCTCCCTCAGGCC 871

RESULT 2  
 US-07-914-281-5/c  
 : Sequence 5, Application US/07914281  
 : Patent No. 5324663  
 GENERAL INFORMATION:  
 APPLICANT: LOWE, JOHN B.  
 TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
 TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
 OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES  
 NUMBER OF SEQUENCES: 14  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: P.O. BOX 1755 JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR  
 CITY: ARLINGTON  
 STATE: VIRGINIA  
 COUNTRY: U.S.A.  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY DISK  
 COMPUTER: IBM PC COMPATIBLE  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentln Release #1.0, Version #1.25

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US/07/914,281
3 FILING DATE: 19920720
4 CLASSIFICATION: 530
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Lavalleye, Jean-Paul M. P.
7 REGISTRATION NUMBER: 31,451
8 REFERENCE/DOCKET NUMBER: 2363-060-55
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (703)521-4500
11 TELEFAX: (703)486-2347
12 TELECY: 248855 OPAT UR
13 INFORMATION FOR SEQ ID NO: 5:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 8174 base pairs
16 TYPE: NUCLEIC ACID
17 STRANDEDNESS: unknown
18 TOPOLOGY: unknown
19 MOLECULE TYPE: DNA (genomic)
20 ANTI-SENSE: NO
21 US-07-914-281-5

```

Query Match	14.68;	Score 184.6;	DB 1;	Length 8174;
Best Local Similarity	78.28;	Pred. No. 9.2e-38;		
Matches 272;	Conservative 0;	Mismatches 64;	Indels 12;	Gaps 4;

OY	795	ATTGAGAAATCTGTTCCTCCAGGCCCTCTCCTTCAAGGGCCACGAGATCCAGGCCCCCAGC	854
Db	1025	ACTTAGGAGTGCAGGCTCGCGGGCCCCCTCTCTCTCTAGACCACGAGATCCAAAGCCCCCTGC	966
OY	855	CCCTCTCTCCTCAAACCAAGG-GTACAGATCCCCAGCCCTCTCTCTCTAGACCACGAG	913
Db	965	CCCTCTCTCTCTCAAGACCCAGAGATCCAGGACCCCAACCCCTCTCTCTCACACCCACGAG	906
OY	914	TCCAGACCCCCCAGCCCTCTCTCTCTAGACCAGAGTCCAG-----CCCTCTCTC	965
Db	905	TCCAGA-TCCCTAGCCCTCTACTCTCTCTAGACCACGAGATCCAGACCAAAAGCTCCCTCTC	847
OY	966	CNTAGAACCCAGAGTCCAGACCCCCCAGCCCTCTCTCTCTAGACCACGAGGTTGAAGC	1025
Db	846	CCTCAGACCCAGGAGCCCAAGTTCCCCAAGCCCTCTCTCTAGATCCAGGAGTTCAGAG-787	
OY	1026	CCCCAACCCCTCTCTCTCTCTAGAGTCCAGAGTCCAAAGCCCAACCCCTCTCTCCCAAC	1085
Db	787	-CCAGAACCCCTCTCTCTCTCTAGACCCAGAGTCCAGGCCCCCAACCCCTCTCTCTCTCAAC	729
OY	1086	CCAAGAGTNNAGTCCAGCCCTCTCTCCNTCAGACCCACGANGTCCAA1133	
Db	728	CCAGAGATCCAGAGCCCCCAGCCCTCTCTCTCTCTAGACCAAGAAAGGCTTA681	

3  
 US-08-393-246-5/C  
 Sequence 5, Application US/08393246  
 Patent No. 5593500  
 GENERAL INFORMATION:  
 APPLICANT: LOWE, JOHN B.  
 TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS  
 OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,  
 TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION  
 OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU  
 TITLE OF INVENTION: 14  
 NUMBER OF SEQUENCES:  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ADDRESSEE: P. C.  
 STREET: 1755 Jefferson Davis Highway, Fourth floor  
 CITY: Arlington  
 STATE: Virginia  
 COUNTRY: U.S.A.  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk

```

STREET: 1755 Jefferson Davis Highway, Fourth Floor
City: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,058A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 8174 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
IS-08-525-058A-5

```

```

Query Match      14.6%; Score 184.6; Db 3; Length 8174;
Best Local Similarity 78.2%; Pred. No. 9,2e-38;
Matches 272; Conservative 0; Mismatches 64; Indels 12; Gaps 4;

QY 795 ATTCAGAGATATCTGTCCAGGCCCTCTCTCCTCAGGCCAGAGTCCAGGCCCCAGC 854
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1025 ACTTAGAGATCGACGGCTCCGGGCCCTCTCTCTCAGACCCAGAGTCCAGGCCCTGC 966

QY 855 CCCTCTCTCCTCAAAACCAAGG-GTACAGATGCCAGGCCCTCTCTCCTCAGACCAGAG 913
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 965 CCTCTCTCTCAGACCCAGAGTCCAGACCCAGGCCCTCTCTCTCAGACCCAGAG 906

QY 914 TCCAGACCCCCCAGCCCCCTCTCTCTCAGACCCAGAGTCCAG-----CCCTCTC 965
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 905 TCCAGA-TCCCTAGCCCCCTACTCTCCTCAGACCCAGAGTCCAGACCAAGCTCTCTC 847

QY 966 CNTCAGACCCAGAGGTCCACACCCCCAGCCCTCTCTCTCTCAGACCCAGAGGTTAGGC 1025
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 846 CCTCAGACCCAGAGGCCAAGTCCCGAGCCCTCTCTCTCAGATCCAGAGTACAGG- 787

QY 1026 CCCCACCCCTCTCTCTCTCAGAGTCCAGAGGTCCAGGCCCCCAACCCCTCTGCCAGAG 1085
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 787 -CCAGACCTCTCTCTCTCAGACCCAGAGTCCAGAGGCCCCCAACCCCTCTCTCTCAGAC 729

QY 1086 CCAGAGGTNNAGATGCCAGCCCTCTCTCTCCTCAGACCCAGAGGTGCCAA 1133
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 728 CCAGAGATCCAGAGGCCCCAGCCCTCTCTCTCCTCAGACACAGAAAGCCCTA 681

RESULT 5
PCT-US91-00899-3/C
: Sequence 3, Application PC/TUS9100899
:
GENERAL INFORMATION:
:
APPLICANT: Lowe, John B.
:
TITLE OF INVENTION: Method and Products For the Synthesis of
:
TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,
:
TITLE OF INVENTION: or as Free Molecules, and For the Isolation of Cloned
:
TITLE OF INVENTION: Genetic Sequences That Determine These Structures
:
NUMBER OF SEQUENCES: 16
:
CORRESPONDENCE ADDRESSES:
:

```

```

: ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
: ADDRESSEE: P.C.
: STREET: 1755 Jefferson Davis Highway, Suite 400
: City: Arlington
: STATE: Virginia
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US91/00899
: FILING DATE: 19910214
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Lavalleye Ph.D., Jean-Paul
: REGISTRATION NUMBER: 31,451
: REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)521-5940
: TELEFAX: (703)486-2347
: TELEX: 24885 OPAT UR
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 8174 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: TISSUE TYPE: Blood
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 4686..5780
: OTHER INFORMATION: /label= mat_peptide
PCT-US91-00899-3
```

```

Query Match      14.6%; Score 184.6; DB 5; Length 8174;
Best Local Similarity 78.2%; Pred. No. 9,2e-38;
Matches 272; Conservative 0; Mismatches 64; Indels 12; Gaps 4;

QY 795 ATTAGGAATATGTGTCCAGCCCTCTCCCTCAGAGCCAGAGATCCAGAGCCAGC 854
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1025 ACTTAGAGTCCAGGCTCCGCCCCCTCTCTCTCAGACCCAGAGTCCAGAGCCCTTGC 966

QY 855 CCTCTCTCCCTCAACAAGG-GTACAGATCCAGCCCTCTCTCTCAGAGCCAGAG 913
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 965 CCTCTCTCTCTCAACAAGGAGTCCAGAGCCAGCCCTCTCTCTCAGAGCCAGAG 906

QY 914 TCCAGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 965
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 905 TCCAGA-TCCCTACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 847

QY 966 CNTAGAGCCAGAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 846 CCTAGAGCCAGAGCCAGAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 787

QY 1026 CCCCACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1085
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 787 -CCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 729

QY 1086 CCAGAGTNNAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1133
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 728 CCAGAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 681
```

RESULT 6  
US-08-308-949A-1/c  
; Sequence 1, Application US/08308949A

```

: Patent No. 5580703
: GENERAL INFORMATION:
: APPLICANT: Kotlin, Robert M.
: APPLICANT: Berns, Kenneth I.
: APPLICANT: Linden, Ralph M.
: TITLE OF INVENTION: Human Adeno-Associated Virus Integration
: TITLE OF INVENTION: Site DNA and Uses Thereof
: NUMBER OF SEQUENCES: 23
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millitia Drive
: CITY: Lexington
: STATE: MA
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/308,949A
: FILING DATE: September 20, 1994
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/947,127
: FILING DATE: September 27, 1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Carroll, Alice O.
: REGISTRATION NUMBER: 33,542
: REFERENCE/DOCKET NUMBER: ACC92-10F
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4060 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
US-08-308-949A-1
```

```

Query Match      11.5%; Score 145.2; DB 1; Length 4060;
Best Local Similarity 72.5%; Pred. No. 6,2e-28;
Matches 243; Conservative 0; Mismatches 78; Indels 14; Gaps 4;

QY 813 CCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 872
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 4000 CCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3941

QY 873 AGGATACAGATCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 932
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3940 GGAATCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3884

QY 933 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 984
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3883 CCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3824

QY 985 GACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1044
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3823 G--GCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3766

QY 1045 AGAGTACAGAGTCCAGAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1104
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3765 AGAGTACAGAGTCC-AGGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3707

QY 1105 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1139
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3706 CCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3672
```



```

: APPLICANT: Little, Sheila P.
: TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
: TITLE OF INVENTION: RELATED NUCLEIC ACIDS
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: United States of America
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US96/04294
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/416,257
: FILING DATE: 04-APR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Blalock, Donna K.
: REGISTRATION NUMBER: 38,082
: REFERENCE/DOCKET NUMBER: X9239
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-277-1090
: TELEFAX: 317-276-3861
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1089 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: mRNA
: PCT-US96-04294-3

Query Match          7.6%; Score 96.6; DB 5; Length 1089;
Best Local Similarity 59.8%; Pred. No. 7.3e-16;
Matches 162; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

QY 457 TGCCACCGTGGCGAGCGTGAACGTCGCGGTGTGTGAGAGGTCGACGTAAGC 516
DB 498 TTCCCTCTACCTCATGTGCGTGATGTCAAGCTCATCTCCCCCGAGCTGCGAGAGG 439
QY 517 TGTATGACCCGCTGTACACACCCGACATGTTCGCGCGCGGAGGCAAGACAGAG 576
DB 438 TTACAAAGACTTACTGAAAAATTCATGTGTGCGCTGGCATCCCGACTCCAGAAAAA 379
QY 577 ACTCTGCAACGGTACTGTGGGGGCCCTGTATCTGCAACGGGTACTGCAAGGCTTG 636
DB 378 ACGCCCTGCAATGTGTACTAGGGGACCGTGTGTGTGCAAGAGTACCTCGTAGGTCTG 319
QY 637 TGTCTTGGAAAAAGCCCGTGTGCGCAAGTTGGCGTGCAGGTGCTACACCACTCT 696
DB 318 TGTCTGGGGAACTTCTCCCTTGGGGCAACCCATATACCGAGAGTCTACACTCAAGTCT 259
QY 697 GCAAAATTCAGTAGGATAGAGAAAAACCGT 727
DB 258 GCAAGTTACCAAGTAGATAAATGACACCAT 228

RESULT 10
: US-08-557-146-1
: Sequence 1, Application US/08557146
: Patent No. 5834290
: GENERAL INFORMATION:
: APPLICANT: Egellrud, Torbjorn
: APPLICANT: Hansson, Lennart
: TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
```

```

: TITLE OF INVENTION: Enzyme (SCCE)
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: White & Case, Patent Department
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2787
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/557,146
: FILING DATE: 14-DEC-1995
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Sterner, Richard J.
: REGISTRATION NUMBER: 35,372
: REFERENCE/DOCKET NUMBER: 1103326-181
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 819-8783
: TELEFAX: (212) 354-8113
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 986 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 25..786
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 25..90
: FEATURE:
: NAME/KEY: mat_peptide
: LOCATION: 112..783
: US-08-557-146-1

Query Match          7.5%; Score 95; DB 4; Length 986;
Best Local Similarity 59.4%; Pred. No. 1.8e-15;
Matches 161; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 457 TGCCACCGTGGCGAGCGTGAACGTCGCGGTGTGTGAGAGGTCGACGTAAGC 516
DB 500 TTCCCTGTACCTCATGTGCGTGATGTCAAGCTCATCTCCCCCGAGCTGCGAGAG 559
QY 517 TGTATGACCCGCTGTACACACCCGACATGTTCGCGCGCGGAGGCAAGACGAAAG 576
DB 560 TTTCAGAGACTTACTGAAAAATTCATGTGTGCGTGTGCAATCCCGACTCCAGAAAA 619
QY 577 ACTCTGCAACGGTACTGTGGGGGCCCTGTATCTGCAACGGGTACTGCAAGGCTTG 636
DB 620 ACGCTGTCAATGTGTACTAGGGGACCGTTGTGTGTGCAAGAGTACCTGCAAGGTCTG 679
QY 637 TGTCTTGGAAAAAGCCCGTGTGCGCAAGTTGGCGTCCAGAGTCTACACCAACTCT 696
DB 680 TGTCTGGGGAACTTCTCCCTTGGGGCAACCAATGACCCAGAGACTACACTCAAGTGT 739
QY 697 GCAAAATTCAGTAGGATAGAGAAAAACCGT 727
DB 740 GCAAGTTACCAAGTAGATAAATGACACCAT 770
```

## RESULT 11

US-08-790-137-2

; Sequence 2, Application US/08790137

; Patent No. 5840871

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Goll, Surya K.

; TITLE OF INVENTION: A NOVEL PROSTATE-ASSOCIATED

; TITLE OF INVENTION: KALLIKREIN

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/790,137

; FILING DATE: Filed Herewith

; CLASSIFICATION: 424

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0195 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-845-4166

; TELEFAX: 415-855-0555

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 833 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-790-137-2

Query Match 6.98; Score 86.8; DB 4; Length 833;

Best Local Similarity 53.98; Pred. No. 2e-13;

Matches 178; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

```
OY 410 CTGCCAGTGGCGGGGGGTGACCCAGAGCTGCTGGTCCAGCAGCAATGCTTACCGTCT 469
    |||||
DB 458 CTGCTCATTTTCGGCTGGGGGACAGCTCCAGCCCGAGTTAGCGCTCTCACACCTT 517
    |||||
OY 470 GCAGTGGGTGAACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 529
    |||||
DB 518 GCGATGGGCGCAATCATCATCATCATCATCATCATCATCATCATCATCATCATCAT 577
    |||||
OY 530 GTACACCCCGCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 589
    |||||
DB 578 CATCACAGACACATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 637
    |||||
OY 590 TGAATCTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 649
    |||||
DB 638 TGAATCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 697
    |||||
OY 650 AGCCCGGTGTGGCCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 709
    |||||
DB 698 GGATCCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 757
    |||||
OY 710 GTGATAGAGAAACCGTCCAGGCCAGTTA 739
    |||||
DB 758 CTGATCCAGAGAGATGAAGAAACAATTA 787
    |||||
```

## RESULT 12

US-08-738-413B-8

; Sequence 8, Application US/08738413B

; Patent No. 5821106

; GENERAL INFORMATION:

; APPLICANT: CHUNG, Kwang-Hoe

; APPLICANT: KOH, You-Seok

; APPLICANT: HWANG, Jae-Hoon

; APPLICANT: KIM, Doo-Sik

; APPLICANT: MOON, Hong-Mo

; TITLE OF INVENTION: A NOVEL CDNA OF DIRECT-ACTING

; TITLE OF INVENTION: FIBRINOLYTIC SERINE PROTEASE

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Dady &amp; Dady PC

; STREET: 805 Third Avenue

; CITY: New York

; STATE: New York

; COUNTRY: US

; ZIP: 10022

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/738,413B

; FILING DATE: October 23, 1996

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, S. Peter

; REGISTRATION NUMBER: 25,351

; REFERENCE/DOCKET NUMBER: 0136/0C539

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-527-7700

; TELEFAX: 212-753-6237

; TELEX: 236687

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 699 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA to mRNA

; ORIGINAL SOURCE:

; ORGANISM: Agkistrodon halys brevicaudus

; IMMEDIATE SOURCE:

; CLONE: protease

; US-08-738-413B-8

Query Match 5.88; Score 73.8; DB 3; Length 699;

Best Local Similarity 63.08; Pred. No. 3.5e-10;

Matches 114; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```
OY 540 AGCATGTTCTGCGCGGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 599
    |||||
DB 481 AGAACAATTGTGTGAGAGTATCTGGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 540
    |||||
OY 600 GGGGCGCTGATCTGCAAGGGGTACTTGACAGGCGCTGTGTGTGTGTGTGTGTGTGTGT 659
    |||||
DB 541 GGAACCCCTCATCTGTATGTGACAAATTCAGGCGCATTCATATTGAGGGGCGGATCTTGT 600
    |||||
OY 660 GGGCAAGTTGGCGCGGAGGTGTCTACCAACTCTGCAAAATTCAGTGTGATGAG 719
    |||||
DB 601 GCCCAACCGGTGAGCGCTGCTACACCAAGGCTTTGATTATATGATTGATGCCAA 660
    |||||
OY 720 A 720
    |||||
DB 661 A 661
    |||||
```

## RESULT 13

US-08-684-862-11

Sequence 11, Application US/08684862

Patent No. 5759541

GENERAL INFORMATION:

APPLICANT: Bach, Alfred

APPLICANT: Hiltner, Heinz

APPLICANT: Bialojan, Siegfried

TITLE OF INVENTION: No. 5759541el Proteins, the Preparation and Use

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kell &amp; Melnkau

STREET: 1101 Connecticut Avenue

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb storage

COMPUTER: IBM AT-compatible, 80286 processor

OPERATING SYSTEM: MS-DOS version 5.0

SOFTWARE: Wordperfect version 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/684,862

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/361,705

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/966,040

FILING DATE: 30-DEC-1992

APPLICATION NUMBER: PCT/EP91/01361

FILING DATE: 19-JUL-1991

INFORMATION FOR SEQ ID NO: 11:

SEQUENCE CHARACTERISTICS:

LENGTH: 957 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA to mRNA

ORIGINAL SOURCE:

ORGANISM: Agkistrodon rhodostoma

FEATURE:

LOCATION: 210 to 911

OTHER INFORMATION: the coding region shown in (2)(1x)(B)

OTHER INFORMATION: codes for the protein of SEQ ID NO: 6

US-08-684-862-11

Query Match 5.4%; Score 68.8; DB 3; Length 957;  
Best Local Similarity 59.0%; Pred. No. 7.1e-09;  
Matches 118; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 521 TGACCCCTGTACACCCAGCATGTCTGCGCGGAGGAGGAGCAACAGAGCTC 580  
DB 674 TGAGGTCTGTGCAAGATATGATGTCGAGGTATGTCGCAAGAGGCAAGATAC 733  
QY 581 CTGCAAGGTGACTCTGGGGGCGCCCTGATCTGCAACGGGTACTTGAGGCGCTGTGTC 640  
DB 734 ATGATATATGATCTCTGGAGAGCTCTCATCTGTATGAAACAGTCCAGGCGCATTTATC 793  
QY 641 TTTCGAAAAAGCCCGTGTGCGCAAGTTGGCGTCCAGGTGTCTACACCAACCTCTGCAA 700  
DB 794 TTATGGGGCCATCTGTGCGCAACCTCTTAAGCTGTATATACACCAAGGCGCATGTA 853  
QY 701 ATTACATGATGATAGAGA 720  
DB 854 TTATATGACTGATCAACA 873

## RESULT 14

US-08-232-463-14

Sequence 14, Application US/08232463

Patent No. 5670367

GENERAL INFORMATION:

APPLICANT: DORNER, F.

APPLICANT: SCHEITLINGER, F.

APPLICANT: FALKNER, F. G.

TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS

NUMBER OF SEQUENCES: 52

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 1800 Diagonal Road, Suite 500

CITY: Alexandria

STATE: VA

COUNTRY: USA

ZIP: 22313-0299

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/232,463

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/935,313

FILING DATE:

APPLICATION NUMBER: EP 91 114 300.6

FILING DATE: 26-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.

REGISTRATION NUMBER: 29,768

REFERENCE/DOCKET NUMBER: 30472/114 IMMU

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)836-9300

TELEFAX: (703)683-4109

TELEX: 899149

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 7218 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

IMMEDIATE SOURCE:

CLONE: PTZgpt-F1s

US-08-232-463-14

Query Match 5.2%; Score 66; DB 2; Length 7218;  
Best Local Similarity 2.3%; Pred. No. 6.5e-08;  
Matches 9; Conservative 232; Mismatches 143; Indels 0; Gaps 0;

QY 789 GAGGAATTCAGGAATATGATCTCCAGGCGCCCTCCCTCAGGCGCCAGAGTCCAGGCC 848  
DB 1052 GAGGAGAGCTTGCAAT 1111  
QY 849 CCGAGCCCTCTCCCTCAACCAAGGATACAGATCCAGCCCTCTCCCTCAGAGCC 908  
DB 1112 YY 1171  
QY 909 AGAGATCCAGAGCCCGCCAGCCCTCTCCCTCAGAGCCAGAGTCCAGGCGCTCTCCT 968  
DB 1172 YY 1231  
QY 969 CAGACCCAGAGTCCAGAGCCCGCCAGCCCTCTCCCTCAGAGCCAGAGGCTGAGGCCCC 1028  
DB 1232 YY 1291  
QY 1029 CAGCCCTCTCTCTGAGATGAGAGTCCAGAGCCCGCCAGCCCTCTCTCCAGAGCCA 1088



Search completed: September 28, 1999, 11:33:30  
Job time: 2090 sec

Db	1282	YY	1351
Qy	1089	GAGGTNNAGGCGCCAGGCCCTCTTCCTCAGACCGNGSTGCATTCGCACCTAGATTTT	1148
Db	1352	YY	1411
Qy	1149	CCCTGNNACACAGTCGCCCTTGTC	1172
Db	1412	YYYYYYYYYYYYYYYYYYYYYYYYGG	1435

RESULT 15  
 US-08-467-155A-2  
 Sequence 2, Application US/08467155A  
 Patent No. 5736377  
 GENERAL INFORMATION:  
 APPLICANT: Band, Ymla  
 TITLE OF INVENTION: NRS-1 POLYPEPTIDES, DNA, AND RELATED  
 TITLE OF INVENTION: MOLECULES AND METHODS  
 NUMBER OF SEQUENCES: 11  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fish & Richardson P.C.  
 STREET: 225 Franklin Street  
 CITY: Boston  
 STATE: MA  
 COUNTRY: USA  
 ZIP: 02110-2804  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/467,155A  
 FILING DATE: 06-JUN-1995  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Clark, Paul T.  
 REGISTRATION NUMBER: 30,162  
 REFERENCE/DOCKET NUMBER: 00398/100001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617/542-5070  
 TELEFAX: 617/542-8906  
 TELEX: 200154  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1454 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: CDNA  
 US-08-467-155A-2

	Query Match	5.2%	Score 65.4	DB 3	Length 1454	
	Best Local Similarity	54.9%	Pred. No. 5.7e-08			
	Matches 129	Conservative	0	Mismatches 106	Indels	Gaps
QY	511	GTAACTCTATGACCCCGGTGACACCCCGAGATGTTCTGCCCGCGGAGGCGAAGACC	570			
Db	683	GTGAGGTCTTTCACCTGGCGGTGCACCAACACATGATATGCTGGACTGGAACGGG	742			
QY	571	AGAAAGACTCTCGCAACGGTGACTCTGGGGGGCCCTATCTGCAACGGGTACTTGCAG	630			
Db	743	GCCAGAGACCCCTTGCCAGATGACTGTGGAGGCCCTCGTGTGTCGAGACCCCTTCAAG	802			
QY	631	GCCTTGTGTTCTTGGAAAAAGCCGCTGGGCCCAAGTTGGCTGCGCAGGCTGTCTACCA	690			
Db	803	GCATCCTCTCGGGGGTGTTCACCCCTGTGGGCTCTGCCAGCATCAGGTGTCTAACCC	862			
QY	691	ACCTTGCAATTCACCTAGTGGATAGAAAAACGTCGAGGCGAGTTAACTCTG	745			
Db	863	AGATGTGCAAAATCATGTCTCTGGATCAATTAAGTATATAGCTTCACATATCCAG	917			

---





F:48,103,196/Active site: His, Asp, Ser #status predicted

alignment\_scores:                      Length:    253  
                     Quality:    459.00                      Gaps:    6  
                     Ratio:    2.765                      Percent Similarity:    65.613                      Percent Identity:    39.130

alignment\_block:  
 US-09-030-606-173 x A44284    ..

Align seg 1/1 to: A44284 from: 1 to: 244

```

10  CACTGCGAGCCCTGCGAGCGGCGACATGCTGATGGAAGAAATGTTCTG 59
   |||||.....|.....|.....|.....|.....|.....|.....|
19  AsnserglnProTyrGlnValAlaIleIle.....AsnGluTyrLeuGly 33
   |||||.....|.....|.....|.....|.....|.....|.....|
60  CTCGGGCGCTGCTGTCATCCGCGAGTGGGCTGTCAGCCCGACACTGT 109
   |||||.....|.....|.....|.....|.....|.....|.....|
33  sGlyGlyValIleuIleAspProSerTyrValIleThrAlaIleHisCys 50
   |||||.....|.....|.....|.....|.....|.....|.....|
110 TCAGAGCTCTTACACCATCGGGGCTGGCGCGACATGCTGAGCGCGAC 159
   |||||.....|.....|.....|.....|.....|.....|.....|
50  YrSerAsnTyrTyrHisValIleuLeuGlyArgAsnAsnLeuPheGluAsp 66
   |||||.....|.....|.....|.....|.....|.....|.....|
160 CAGAGCGGAGCGGACGACATGCTGGAGCGCGACCTCTCCGACGCGACC 209
   |||||.....|.....|.....|.....|.....|.....|.....|
67  ..GluProPheAlaGlnTyrArgPheValAsnGlnSerPheProHisPsr 82
   |||||.....|.....|.....|.....|.....|.....|.....|
210 AGACTACACAGACCCCTGCTC..... 231
   |||||.....|.....|.....|.....|.....|.....|.....|
82  cAspTyr...LysProPheLeuMetArgAsnHisThrArgGlnArgGlyA 98
   |||||.....|.....|.....|.....|.....|.....|.....|
232  ....GCTAAGCACTCATGCTGCATCAGATGCGAGATCCGCTGCC 273
   |||||.....|.....|.....|.....|.....|.....|.....|
98  sPsrAspTyrSerAsnAspLeuMetLeuIleHisLeuSerGlnProAlaAsp 114
   |||||.....|.....|.....|.....|.....|.....|.....|
274  GAGTGTGACACCATCCGAGCATCAGCATGCTTCCGAGTCCCTACCGCGC 323
   |||||.....|.....|.....|.....|.....|.....|.....|
115  IleThrAspGlyValIleValIleAspLeuProThrGlnGluProGlyVal 131
   |||||.....|.....|.....|.....|.....|.....|.....|
324  GGGGAACTCTTGCTGCTTGTGCTGGGCTGCTGGCGAAGCGGTGAGC 373
   |||||.....|.....|.....|.....|.....|.....|.....|
131  IGlYserThrCysLeuAlaSerGlyTyrGly..... 141
   |||||.....|.....|.....|.....|.....|.....|.....|
374  TCAGGGGTGTGTCTGCCCTTTCAGAGAGTCTCTGCCACAGCGGG 423
   |||||.....|.....|.....|.....|.....|.....|.....|
142  .... 144
   |||||.....|.....|.....|.....|.....|.....|.....|
424  GGGGTGACCA..GAGCTGTGCTGCCAGCAAGATGCTACCGTGTGCA 472
   |||||.....|.....|.....|.....|.....|.....|.....|
145  ProLeuAsnTyrGlnLeu.....ProAspAspLeuGln 155
   |||||.....|.....|.....|.....|.....|.....|.....|
473  GTGGGTGAAGCTGTGCTGTGTGTGAGAGAGTGTGACAGTAACTCTTAC 522
   |||||.....|.....|.....|.....|.....|.....|.....|
155  nCyValAsnIleHisLeuSerAsnGlnIleCysGlyIleGlnAlaTyrG 172
   |||||.....|.....|.....|.....|.....|.....|.....|
523  ACCCGGTGTACACCCCGCATGTTCTGCCGCGGCGAGGCGACAGCAG 572
   |||||.....|.....|.....|.....|.....|.....|.....|
172  IuGlnIleValThrAspLeuMetLeuCysAlaGlyAlaMetAspIleArg 188
   |||||.....|.....|.....|.....|.....|.....|.....|
573  AAGGACTCTGCAAGGCTGACTGTGGGGGCGCGGATGTGCAACAGGGTA 622
   |||||.....|.....|.....|.....|.....|.....|.....|
189  LysAspThrCysLysGlyAspSerGlyIleProLeuIleCysAspGlyVal 205
   |||||.....|.....|.....|.....|.....|.....|.....|
623  CTTCGAGGCGCTGTGCTTTCGAAAAAGCCCGTGTGGCAAGTTGGCG 672
   |||||.....|.....|.....|.....|.....|.....|.....|
205  IleuGlnIleIleThrSerTyrGlyAsnValProCysAlaGlnProTyrA 222
   |||||.....|.....|.....|.....|.....|.....|.....|
673  TGCCAGGTGTACACCACTCTGCAAAATTCATGAGTGGATAGAGAAA 722
   |||||.....|.....|.....|.....|.....|.....|.....|
222  snProGlyValIleThrIleHisLeuIleLysPheThrSerTyrIleLysGlu 238
   |||||.....|.....|.....|.....|.....|.....|.....|

```

723 ACCGTCGAG 731  
 ..|||.....  
 239 ValMetLys 241

seq\_name: p1r1:TRDC

seq\_documentation\_block:

trypsin (EC 3.4.21.4) precursor, cationic - dog  
 N:Alternate names: cationic trypsinogen  
 C:Species: Canis lupus familiaris (dog)  
 C:Date: 30-Sep-1987 #sequence-revision 30-Sep-1987 #text-change 22-Jan-1999  
 C:Accession: B26273  
 R:Pinsky, S.D.; LaForge, K.S.; Scheele, G.  
 Mol. Cell. Biol. 5, 2669-2676, 1985  
 A:Title: Differential regulation of trypsinogen mRNA translation: full-length mRNA se  
 A:Reference number: A26273; MVID:86284628  
 A:Accession: B26273  
 A:Molecule type: mRNA  
 A:Residues: 1-246 <PIN>  
 A:Cross-references: GB:M11590; NID:G164096; PID:G164097  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; pancreas; protein digestion; serine protease; zymogen  
 F:1-15/Domain: signal sequence #status predicted <SIG>  
 F:16-23/Domain: activation peptide #status predicted <APR>  
 F:24-246/Product: trypsin, cationic #status predicted <ENZ>  
 F:24-239/Domain: trypsin homology <TRY>  
 F:30-160,48-64,132-233,139-206,171-185/Disulfide bonds: #status predicted  
 F:53,107,200/Active site: His, Asp, Ser #status predicted  
 F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

alignment\_scores:                      Length:    245  
                     Quality:    457.00                      Gaps:    6  
                     Ratio:    2.737                      Percent Similarity:    68.163                      Percent Identity:    38.776

alignment\_block:

US-09-030-606-173 x TRDC

Align seg 1/1 to: TRDC from: 1 to: 246

```

4  AGCCGCACTGCGAGCCCTGGCGAGCGGCGACATGCTGATGGAAGAAAT 53
   |||.....|.....|.....|.....|.....|.....|.....|
31  SerArgAsnSerValProTyrGlnValSerLeu...AsnserGlyTyrH 46
   |||||.....|.....|.....|.....|.....|.....|.....|
54  GTTGTGCTGGGCGCTCTGTGTCATCCGCGAGTGGGCTGTGCAAGCCGAC 103
   |||||.....|.....|.....|.....|.....|.....|.....|
46  sPheCysGlyGlySerLeuIleAsnserGlnTyrValIleSerAlaIle 153
   |||||.....|.....|.....|.....|.....|.....|.....|
104  ACTGTTCCGAACTCTTACACCATGCGGGCTGGCGCGCGACATGCTGAG 153
   |||||.....|.....|.....|.....|.....|.....|.....|
63  IsCysTyrLysSerArgIleGlnValArgLeuGlyGluTyrAsnIle... 78
   |||||.....|.....|.....|.....|.....|.....|.....|
154  GCCGACCAAGAGCCAGGAGGAGCCAGATGTGGAGGCCACTCTCCGTACG 203
   |||.....|.....|.....|.....|.....|.....|.....|
79  AlaValSerGlnGlyGlyGlnGlnPheIleAsnAlaIleLysIleIleArg 95
   |||||.....|.....|.....|.....|.....|.....|.....|
204  GCACCCAGAGTACACAGACCCCTTGTGCTACAGCACTCATGCTATCA 253
   |||||.....|.....|.....|.....|.....|.....|.....|
95  gHisProArgTyrAsnAlaAsnThrIleAspAsnAspIleMetLeuIle 112
   |||||.....|.....|.....|.....|.....|.....|.....|
254  AGTGGAGCAATCGTGTGCGAGTGTGACACCATCCGAGCATGAGATT 303
   |||||.....|.....|.....|.....|.....|.....|.....|
112  YrLeuSerSerProAlaThrLeuAsnSerArgValSerAlaIleAlaLeu 128
   |||||.....|.....|.....|.....|.....|.....|.....|
304  GCTTCGAGTCCCTTACCGCGGCGGAACTTGTGCTCTTGTGCTGGG 353
   |||||.....|.....|.....|.....|.....|.....|.....|
129  ProLysSerCysProAlaIleGlyThrGlnCysLeuIleSerGlyTyrPgl 145
   |||||.....|.....|.....|.....|.....|.....|.....|
354  TCTGTGCGGACGCGTGAAGCTCAGCGGCTGTGTCTGCTCCCTTCAAGA 403
   |||||.....|.....|.....|.....|.....|.....|.....|
145  Y..... 145

```

```

404 GGTCTCTGCCAGTCGGGGGCTGACCCAGACCTCTGCTCCAGCA 453
      ::::::::::::::|::|
146 .....AsnThrGlnSerIleGly.....Gln..A 153
454 GAATGCTTACCGTCTGACAGTGCCTGAACGTCGTGGTCTCGAGAG 503
      ::|::|::|::|::|::|::|::|::|::|::|::|::|
153 snYrProAspValLeuGlnCysLeuYsaIaProlIleLeuSerAspSer 169
504 GTCGTGACAGCTCTATGACCCGCTGACCCAGACCTGTTCTCTCGC 553
      |||||::|::|::|::|::|::|::|::|::|
170 ValCysArgAsnAlaTyrrProGlyGlnIleSerSerAsnMetMetCysLe 186
554 CGCGGAGGCGCAGACCCAGAGACCTCTGCACAGCTGCTGGGGGC 603
      |||::|::|::|::|::|::|::|::|::|
186 uGlyTyrMetGlnCylGlyLysAspSerCysGlnGlyAspSerGlyLyp 203
604 CCTGATCTCAACGGGTACTGTGACGGGCTGTGTCTTCGGAAAGCC 653
      ||::|::|::|::|::|::|::|::|::|
203 roValValCysAsnGlyLeuLeuGlnIlyValValSerTrpGly...Ala 218
654 CCGTGTGGCAAGTGTGGCTGCGAGGTGTCTACACAACTCTCAAT 703
      ||::|::|::|::|::|::|::|::|::|
219 GlyCysAlaGlnLysGlyLysProGlyValSerProLysValCysLys 235
704 CACTGAGTGAATAGAGAAACCTCCAGGCCAGT 737
      ::::::::::::::|::|
235 rValSerTrpIleGlnGlnThrIleAlaAlaAsn 246
seq_name: p1r2:S45303

```

```

seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) precursor - dog
  C:Species: Canis lupus familiaris (dog)
  C>Date: 20-Oct-1994 #sequence_revision 03-Aug-1995 #text_change 17-Mar-1999
  C:Accession: S45303; S38487
  R:Gauthier, E.R.; Dumas, C.; Chappelain, P.; Tremblay, R.R.; Dube, J.Y.
  Biochim. Biophys. Acta 1218, 102-104, 1994
  A:Title: Characterization of canine pancreas kallikrein cDNA.
  A:Reference number: S45303; MUID:94250683
  A:Accession: S45303
  A:Molecule type: mRNA
  A:Residues: 1-261 <GAP>
  A:Cross-references: EMBL:X75479; NID:g414018; PID:g414019
  C:Superfamily: trypsin; trypsin homology
  C:Keywords: hydrolase; serine proteinase
  F:1-24/Domain: signal sequence #status predicted <STG>
  F:25-261/Product: tissue kallikrein #status predicted <MAT>
  F:25-253/Domain: trypsin homology <TRI>
  F:65,120,213/Active site: His, Asp, Ser #status predicted

```

```

alignment_scores:
  Quality: 456.50      Length: 254
  Ratio: 2.750        Gaps: 5
  Percent Similarity: 65.354      Percent Identity: 37.008

```

```

alignment_block:
  US-09-030-606-173 x S45303 ..

```

```

Align seg 1/1 to: S45303 from: 1 to: 261

```

```

10 CACTGCGACGCTGCGAGCGGCGCTGTCATGGAACGAATTTGTCG 59
   ::::::::::::::|::|
34 AsnSerGlnProTrpGlnAlaIleLeuTyrHisTyrSerLysPheGlnCys 50
60 CTGGGGGCTGCTGTCGATCGGAGTGGGTGTCTACGGCGACACTGT 109
   ::::::::::::::|::|
50 scGlyGlyValLeuValHisProGlyTrpValValThrAlaAlaHisCysI 67
110 TCAGAACTCTTACACATCGGGCTGGGCTGCACACTGTGAGCGCGAC 159
   ::::::::::::::|::|
67 LeuAsnAspAsnTyrGlnLeuTrpLeuGlyArgTyrAsnLeu...PheGln 82

```

```

160 CAAGACCCAGGAGCCAGATGTGAGAGCCAGCTCTCCGTCAGCCACC 209
   ::|::|::|::|::|::|::|::|::|
83 HisGluAspThrAlaGlnPheValGlnValArgGluSerPheProHisPsr 99
210 AGAGTCAACACAGACCTTGCTGCTTAC..... 237
   ::|::|::|::|::|::|::|::|::|
99 OGluPheAsnLeuSerLeuLysAsnHisThrArgLeuProGluGluA 116
238 .....GACCATGCTCATCAAGTGTGAGAGAACCTCGGTCCGAG 276
116 spTYrSerHisAspIleMetLeuLeuArgLeuAlaGluProAlaGlnIle 132
277 TCTGACACCATCCGAGCATCAACATGCTTCTGCACTGCTCACCGCGG 326
   ::|::|::|::|::|::|::|::|::|
133 ThrAspAlaValArgValLeuAspLeuProThrGlnGluProGlnValAl 149
327 GAACCTTGCTGCTGTTTGTGCTGGGGTGTCTGCTGGCGAAGGTGACTCA 376
   ::|::|::|::|::|::|::|::|::|
149 ySerThrCysTyrAlaSerGlyTrpGlySerIle...Glu..... 161
377 CGGGTGTGTGTGCGCCCTGTCAAGAGGTCTCTGCCAGTGGCGGGG 426
161 ..... 161
427 CTGACCCAGAGCTCTGCTCCAGGCGAGATGCTTACCTGCTGCACTGC 476
   ::|::|::|::|::|::|::|::|::|
162 .....ProAspLysPheIle.....TyrProAspAspLeuGlnCys 173
477 GTGACAGTGTGCTGCTGTCTGAGAGAGTCTGCAGTACCTATGACCC 526
   ::|::|::|::|::|::|::|::|::|
174 ValAspLeuGlnLeuLeuSerAsnAspIleCysAlaAsnAlaHisSerG 190
527 GCTGTACCAACCCAGCATGTCTGCGCGCGGAGGCGCAAGACCAAG 576
   |||::|::|::|::|::|::|::|::|
190 nLysValThrGluPheMetLeuCysAlaGlyHisLeuGlnGlyGlySA 207
577 ACTCTGCAACGGTGACTCTGGGGGCCCCGTATCTGCAACGGTACTGG 626
   |||::|::|::|::|::|::|::|::|
207 spHrcysValGlyAspSerGlyGlyProLeuIleCysAspGlyValLeu 223
627 CAGGCTGTGTCTTTCGAAAGCCCGGTGTGGCAAGTGGCGTGC 676
   |||::|::|::|::|::|::|::|::|
224 GlnGlyIleThrSerTrpGlyHisValProCysGlySerProAsnMetPr 240
677 AGGTGCTACACCACTCTGCAATTCAGTGAAGATGAGAAACCG 726
   ::|::|::|::|::|::|::|::|::|
240 OAlaValIlyThrLysValIleSerHisLeuGlnTrpIleLysGluThrM 257
727 TCCAGGCCAGT 737
   ::|::|::|
257 etThrAlaAsn 260

```

```

seq_name: p1r2:B31136
seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) 3 precursor, submandibular - rat
  N:Alternate names: glandular prokallikrein 3, submandibular
  C:Species: Rattus norvegicus (Norway rat)
  C>Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 20-Mar-1998
  C:Accession: B31136
  R:Chen, Y.P.; Chao, J.; Chao, L.
  Biochemistry 27, 7189-7196, 1988
  A:Title: Molecular cloning and characterization of two rat renal kallikrein genes.
  A:Reference number: A31136; MUID:89088074
  A:Accession: B31136
  A:Molecule type: DNA
  A:Residues: 1-259 <CHE>
  A:Cross-references: GB:M19648; GB:J02837; NID:g205002; PID:g205004
  A>Note: the authors translated the codon GTC for residue 230 as Cys
  C:Superfamily: trypsin; trypsin homology
  C:Keywords: hydrolase; serine proteinase
  F:1-18/Domain: signal sequence #status predicted <STG>
  F:19-259/Product: tissue prokallikrein 3, submandibular #status predicted <MAT>

```

F:25-251/Domain: trypsin homology <TRY>  
F:63,118,211/Active site: His, Asp, Ser #status predicted

alignment\_scores:  
Quality: 455.00 Length: 254  
Ratio: 2.758 Gaps: 5  
Percent Similarity: 64.961 Percent Identity: 36.614

alignment\_block:  
US-09-030-606-173 x B31136 ..

Align seg 1/1 to: B31136 from: 1 to: 259

```

10 CACTCGCAGCCCTGGCAGGCGGCTGTCATGAAAGCAATGTTCTG 59
   |||||||.....|.....|.....|.....|.....|.....|
34 AsnSerGlnProTyrGlnValAlaValIle.....AsnArgTyrLeuCy 48
60 CTCGGGCTCTGTCATCCGACATGGGTGCTGTACGCCGACACTGT 109
   |||||||.....|.....|.....|.....|.....|.....|
48 sGlyValLeuIleAsnProSerTyrValIleThrAlaIleHisCys 65
110 TCAGAACTCTACACCATCGGGCTGGGCTGACAGTCTTGAGCCGAC 159
   |||||||.....|.....|.....|.....|.....|.....|
65 yTserHisAsnTyrHisValLeuGlnArgAsnAsnLeuPheLysAsp 81
160 CAAGAGCCAGGAGCAGATGGTGGAGCCGACCTCCGATACGCGACCC 209
   |||||||.....|.....|.....|.....|.....|.....|
82 ..GluProPheAlaGlnTyrArgValValAsnGlnSerPheProHisPr 97
210 AGATGACACAGACCCCTGCTC.....|.....|.....|.....| 231
   |||||||.....|.....|.....|.....|.....|.....|
97 oAspTyrAsnProPhePheMetLysAsnHisThrLeuPheProGlyAsp 114
232 .....GCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCCAG 276
   |||||||.....|.....|.....|.....|.....|.....|
114 sPHisSerAsnAspLeuMetLeuLeuHisLeuSerGlnProAlaAspIle 130
277 TCTGACACCATCGGAGCATGCGATGCTTCGAGTGGCCGACGCGGG 326
   |||||||.....|.....|.....|.....|.....|.....|
131 ThrAspGlyValLysValIleAsnProThrGlnGluProLysValGln 147
337 GAACCTCTGCTCTGTTCTGGCTGGGCTGCTGCTGCGAAGCGTAGCTCA 376
   |||||||.....|.....|.....|.....|.....|.....|
147 ySerThrCysLeuAlaSerGlyTyr.....|.....|.....|.....| 155
377 CGGGTGTGTCTGCCCTCTTCAGAGAGTCTCTGCCAGTCCGGGGG 426
   |||||||.....|.....|.....|.....|.....|.....|
155 .....|.....|.....|.....|.....|.....|.....| 155
427 CTGACCCAGAGCTGCGTCCCGAGGAGCATG.....CTACCGTGTCTG 469
   |||||||.....|.....|.....|.....|.....|.....|
156 .....SerSerThrLysProLeuGlnUtrGluPheProAspAsp 169
470 GCAGTGCCTGAACGTCGTGCTGCTGTCGAGAGGTCGTCAGTAAGTCT 519
   |||||||.....|.....|.....|.....|.....|.....|
169 uGlnCysValAsnIleAsnIleLeuSerAsnGlnLysCysIleLysAla 186
520 ATGAGCCCTGTACCAACCCAGACATGTTCTGCGCGCGGAGGAGCAAGAC 569
   |||||||.....|.....|.....|.....|.....|.....|
186 lSerThrMetValThrAspValMetLeuCysAlaGlyGluLeuGlnCyl 202
570 CAGAGGACTCCTGCAAGGTCATCGGGGCGCCCTGATCTGCAAGG 619
   |||||||.....|.....|.....|.....|.....|.....|
203 GlyLysAspThrCysAsnGlyAspSerGlyGlyProLeuLeuCysAsp 219
620 GTACTTCAGGCGCTGTGTCTTTCGAAAAGCCCGTGGCCAAAGTTG 669
   |||||||.....|.....|.....|.....|.....|.....|
219 yValLeuGlnGlyIleThrSerTyrSerValProCysGlyGluThr 236
670 GCGTCCCAAGGTGTACACCAACTCTCCAAATTCAGTAGTGATAG 719
   |||||||.....|.....|.....|.....|.....|.....|
236 snArgProAlaIleTyrThrLysLeuIleLysPheThrSerTyrLeu 252

```

720 AAACCGTCAG 731  
:.....:  
253 GluValMetLys 256

seq\_name: p1r2:A34079

seq\_documentation\_block:

tissue kallikrein (EC 3.4.21.35) P1 precursor - rat  
N:Alternate names: kallikrein-related proteinase k8  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 30-Mar-1990 #sequence\_revision 30-Mar-1990 #text\_change 08-Sep-1997  
C:Accession: A34079; S10700  
R:Brady, J.M.; Wines, D.R.; MacDonald, R.J.  
Biochemistry 28, 5203-5210, 1989  
A:Title: Expression of two kallikrein gene family members in the rat prostate.  
A:Reference number: A34079; M0ID:89352606  
A:Accession: A34079  
A>Status: preliminary  
A:Molecule type: DNA; mRNA  
A:Residues: 1-261 <RA>  
A:Cross-references: GB:M27215; GB:M27216; GB:M27217; NID:g206638; PID:g206640; GB:M27  
A:Experimental source: prostate  
R:Elmoujahed, A.; Gutman, N.; Brillard, M.; Gauthier, F.  
FEBS Lett. 265, 137-140, 1990  
A:Title: Substrate specificity of two kallikrein family gene products isolated from t  
A:Reference number: S10698; M0ID:90306305  
A:Accession: S10700  
A:Molecule type: protein  
A:Residues: 25-43;112-138 <ELM>  
A:Experimental source: submaxillary gland  
A:Note: 125-Lys was also found  
C:Superfamily: trypsin; trypsin homology  
C:Keywords: hydrolase; serine proteinase  
F:25-253/Domain: trypsin homology <TRY>  
F:25-111/Product: tissue kallikrein P1 light chain #status experimental <MAT1>  
F:112-261/Product: tissue kallikrein P1 heavy chain #status experimental <MAT2>  
F:65,120,213/Active site: His, Asp, Ser #status predicted

alignment\_scores:  
Quality: 452.50 Length: 252  
Ratio: 2.846 Gaps: 5  
Percent Similarity: 63.095 Percent Identity: 38.492

alignment\_block:  
US-09-030-606-173 x A34079 ..

Align seg 1/1 to: A34079 from: 1 to: 261

```

10 CACTCGCAGCCCTGGCAGGCGGCTGTCATGAAAGCAATGTTCTG 59
   |||||||.....|.....|.....|.....|.....|.....|
34 AsnSerGlnProTyrGlnValAlaValIleThrHisPheAsnGlnProGlnCy 50
60 CTCGGGCTCTGTCATCCGACATGGGTGCTGTACGCCGACACTGT 109
   |||||||.....|.....|.....|.....|.....|.....|
50 sGlyValLeuIleHisProSerTyrValIleThrAlaIleHisCys 67
110 TCAGAACTCTACACCATCGGGCTGGGCTGACAGTCTTGAGCCGAC 159
   |||||||.....|.....|.....|.....|.....|.....|
67 yTserValAsnTyrGlnValTyrLeuGlnArgAsnAsnLeuLeuGlnAsp 83
160 CAAGAGCCAGGAGCAGATGTCGAGGCGCCCTGCTCCGACGCGACCC 209
   |||||||.....|.....|.....|.....|.....|.....|
84 ..GluProPheAlaGlnHisArgLeuValSerGlnSerPheProHisPr 99
210 AGATGACACAGACCCCTGCTC.....|.....|.....|.....| 231
   |||||||.....|.....|.....|.....|.....|.....|
99 oGlyPheAsnLeuAspIleIleLysAsnHisThrArgLysProGlyAsn 116
232 .....GCTAACGACCTCATGCTCATCAAGTTGGAGCAATCCGTGCCAG 276
   |||||||.....|.....|.....|.....|.....|.....|
116 sPtySerAsnAspLeuMetLeuLeuHisLeuLysThrProAlaAspIle 132

```





A:Accession: A29746  
 A:Molecule type: DNA  
 A:Residues: 1-259 <DR1>  
 A:Experimental source: strain BALB/c, salivary gland  
 A:Note: the authors translated the codon GAC for residue 210 as Gly  
 R:Evans, B.A.; Drinkwater, C.C.; Richards, R.I.  
 J. Biol. Chem. 262, 8027-8034, 1987  
 A:Title: Mouse glandular kallikrein genes: Structure and partial sequence analysis of th  
 A:Reference number: 15260; MUID:87250386  
 A:Accession: 170039  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 70-120 <RES>  
 A:Cross-references: GB:M18618; NID:9198556; PID:9198561  
 A:Accession: 170038  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 17-54 <RES2>  
 A:Cross-references: GB:M18598; NID:9198555; PID:9198560  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase  
 F:25-251/Domain: trypsin homology <TRY>  
 F:65,118,211/Active site: His, Asp, Ser #status predicted

## alignment\_scores:

Quality:	452.00	Length:	251
Ratio:	2.790	Gaps:	4
Percent Similarity:	64.542	Percent Identity:	38.247

## alignment\_block:

US-09-030-606-173 x A29746

Align seg 1/1 to: A29746 from: 1 to: 259

```

10  CACTCGCAGCCCTGCGAGCGGCGGACGTGTCATGGAAGAAATGTTCTG 59
   ::::::::::::::::::::
34  AAsnSerGlnProtrpGlnValAlaValAlaValAlaValAlaValAlaVal 50
   ::::::::::::::::::::
60  CTGCGGCGTCTGTCATCGGAGAGGAGGCTGTCAGCGGCGGACACTGTT 109
   ::::::::::::::::::::
50  seg1GlyValAlaLeuLeuAspArgAsnTrpValLeuThrAlaAlaHisCys 67
   ::::::::::::::::::::
110 TCCAGAACCTTACACCATCGGCGGCTGCGGCTGACAGTGTGAGCGGCG 159
   ::::::::::::::::::::
67  yGlnAspLysIleAsnIleTyrLeuGlyLysAsnLysLeuPheGlnAsp 83
   ::::::::::::::::::::
160 CAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 209
   ::::::::::::::::::::
84  ...GluProSerAlaGlnHisArgLeuValSerLysSerPheProHisTrp 99
   ::::::::::::::::::::
210 AGAGTACAAACAGACCTTG.....CTCG 232
   ::::::::::::::::::::
99  OAspPheAsnMetSerLeuLeuGlnSerValProThrGlyAlaAspLeu 116
   ::::::::::::::::::::
233 CTAACACCTCATGCTCATCAAGTGTGAGCAATCCCTGCGGAGTGTGAC 282
   ::::::::::::::::::::
116 eTAAsnLysPheMetLeuLeuArgLeuSerLysProAlaAspIleThrAsp 132
   ::::::::::::::::::::
283 ACCATCCGAGCATCAGCATGCTGTCAGAGTGCCTACCGCGGAGAACTC 332
   ::::::::::::::::::::
133 ValValLysProLysLeuPheProThrGlnProLysLeuGlnLysSerThr 149
   ::::::::::::::::::::
149 rCysLeuAlaSerGlyTrpLys..... 156
   ::::::::::::::::::::
383 TGTGTGTGCGCTTTCAGAGAGGTCTGTGCCAGTCGCGGCGGCTGAC 432
   ::::::::::::::::::::
156 ..... 156
   ::::::::::::::::::::
433 CAGAGCTGTGCGTCCAGGC.....AGAACTCATACGCTGCTGAGTG 475
   ::::::::::::::::::::

```

```

157 .....SerIleAsnGlnLeuIleTyrGlnAsnProAsnAspLeuGlnCys 171
   ::::::::::::::::::::
476 CGTGAACGTGGGTGGTGGTCTGAGAGAGTCTGCACTAGACCTTATGAC 525
   ::::::::::::::::::::
171 ValSerIleLysLeuHisProAsnGlnValAlaCysValLysAlaHisIle 188
   ::::::::::::::::::::
526 CCCTGTACCAACCCAGCATGTTCTGCGCGCGGAGGAGGAGCAACAGAG 575
   ::::::::::::::::::::
188 euLysValThrAspValMetLeuCysAlaGlyGlnMetAsnLysGlyLys 204
   ::::::::::::::::::::
576 GACTCTGCAACGGTACTCTGCGGGGCCCCGTATCTGCACGGTACTT 625
   ::::::::::::::::::::
205 AspThrCysLysGlyAspSerGlyProLeuIleCysAspGlyValle 221
   ::::::::::::::::::::
626 GCAGGCGCTGTGTCTTTGGGAAAGCCCCGTGTGCGCAAGTTGGCGTGC 675
   ::::::::::::::::::::
221 uGlnGlyIleThrSerTrpGlySerThrProCysGlyLysProAsnAlaP 238
   ::::::::::::::::::::
676 CAGGTGTACACCAACCTCTGCAATTCATGAGTGATAGAGAAAC 725
   ::::::::::::::::::::
238 roAlaIleLysThrLysLeuIleLysPheThrSerTrpLysAspThr 254
   ::::::::::::::::::::
726 GTC 728
   ::::::::::::::::::::
255 Met 255
   ::::::::::::::::::::
seq_name: p1r1:NGMSG
seq_documentation_block:
7S nerve growth factor gamma chain (EC 3.4.21.-) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 18-Dec-1981 #sequence_revision 17-May-1985 #text_change 20-Mar-1998
C:Accession: A91005; A90949; A93510; A92341; A00942; A21093; A22705
R:Evans, B.A.; Richards, R.I.
EMBO J. 4, 133-138, 1985
A:Title: Genes for the alpha and gamma subunits of mouse nerve growth factor are cont
A:Reference number: A91005; MUID:85257431
A:Accession: A91005
A:Molecule type: DNA
A:Residues: 1-261 <EVA>
R:Ullrich, A.; Gray, A.; Wood, W.I.; Hayflick, J.; Seeburg, P.H.
DNA 3, 387-392, 1984
A:Title: Isolation of a cDNA clone coding for the gamma-subunit of mouse nerve growth
A:Reference number: A90949; MUID:85076169
A:Accession: A90949
A:Molecule type: mRNA
A:Residues: 1-261 <UPL>
A:Cross-references: GB:X01389; NID:953373; PID:953374
R:Howles, P.N.; Dickinson, D.P.; DiCaprio, L.L.; Woodworth-Gutal, M.; Gross, K.W.
Nucleic Acids Res. 12, 2791-2805, 1984
A:Title: Use of a cDNA recombinant for the gamma-subunit of mouse nerve growth factor
A:Reference number: A93510; MUID:84169573
A:Accession: A93510
A:Molecule type: mRNA
A:Residues: 127-202, 'E', 204-261 <HOW>
A:Cross-references: GB:X00472; NID:954260; PID:954261
R:Thomas, K.A.; Baglan, N.C.; Bradshaw, R.A.
J. Biol. Chem. 256, 9156-9166, 1981
A:Title: The amino acid sequence of the gamma-subunit of mouse submaxillary gland 7 S
A:Reference number: A92341; MUID:81264363
A:Accession: A92341
A:Molecule type: protein
A:Residues: 25-107, 112-261 <THO>
A:Experimental source: outbred strain Swiss Webster
C:Comment: 7S nerve growth factor is composed of two alpha chains, a beta dimer compo
C:Genetics: The active form of the gamma chain occurs naturally as combinations of elt
A:Map position: 7
A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; growth factor; hydrolase; serine proteinase; submandibular
F:1-18/Domain: signal sequence #status predicted <SIG>

```

F:25-253/Domain: trypsin homology <TRY>  
 F:25-107/Domain: segment B1 <GB1>  
 F:25-107.112-261/Product: nerve growth factor gamma chain (active form) #status experim  
 F:112-261/Domain: segment A <GAA>  
 F:112-164/Domain: segment C <GCC>  
 F:165-261/Domain: segment B2 <GB2>  
 F:31-173.50-66.152-219.184-198.209-234/Disulfide bonds: #status predicted  
 F:65.120.213/Active site: His, Asp, Ser #status predicted  
 F:102/Binding site: carbohydrate (Asn) (covalent) #status experimental

alignment\_scores:  
 Quality: 450.00 Length: 258  
 Ratio: 2.830 Gaps: 4  
 Percent Similarity: 61.628 Percent Identity: 35.659

alignment\_block:

US-09-030-606-173 x NGMSG ..

Align seg 1/1 to: NGMSG from: 1 to: 261

```

10 CACTGGCAGCCCTGGCAGGCGGCATGTCATGGAAGAAATTTGTTCTG 59
   |||||||.....:.....:.....:.....:.....:.....:
34 AsnSerGlnProThrPheValAlaValAlaValArgTyrTrpGlnTyrLeuCy 50
   CTGGGCGCTCTGTCATCCGACAGTGGGTGCTGCACCGCACACTGTT 109
   |||||||.....:.....:.....:.....:.....:.....:
50 sGlyGlyValLeuLeuAspProAsnTrpValLeuThrAlaAlaHisCysT 67
   |||||||.....:.....:.....:.....:.....:.....:
110 TCACGACTCCTACACCATCGGGCTGGCGCTGCACAGTCTTGAGCCGAC 159
   |||||||.....:.....:.....:.....:.....:.....:
67 YrAspAsnAsnTyrLysValTrpLeuGlyLysAsnAsnLeuPheLysAsp 83
   |||||||.....:.....:.....:.....:.....:.....:
160 CAAGGCCAGGAGGAGATGATGAGGCCAGCCCTCCGTAAGCCGACCC 209
   |||||||.....:.....:.....:.....:.....:.....:
84 ..GluProSerAlaGlnHisArgPheValSerLysAlaAlaIleProHisPr 99
   |||||||.....:.....:.....:.....:.....:.....:
210 AGAGTACAACAGACCCCTGCTC.....:.....:.....:.....: 231
   |||||||.....:.....:.....:.....:.....:.....:
99 OGlyPheAsnMetSerLeuMetArgLysHisIleArgPheLeuGluTyrA 116
   |||||||.....:.....:.....:.....:.....:.....:
232 .....GCTAACACCTCATGCTCATCAAGTTGAGCAATCCGTGTCGAG 276
   |||||||.....:.....:.....:.....:.....:.....:
116 sPtySerAsnAspLeuMetLeuArgLeuSerLysProAlaAspIle 132
   |||||||.....:.....:.....:.....:.....:.....:
277 TCTGACACCATCCGAGCATCAGCATTCGCTGCGAGTCCCTACCGCGGG 326
   |||||||.....:.....:.....:.....:.....:.....:
133 ThrAspTrpValLysProIleThrLeuProTrpGluProLysLeuGl 149
   |||||||.....:.....:.....:.....:.....:.....:
337 GAACCTGCTGCTGTTCTGGCTGGGTCTGCTGGCGAAGGCTGACTCA 376
   |||||||.....:.....:.....:.....:.....:.....:
149 ySerThrLysLeuAlaSerGlyTrpGlySerIleThr.....:.....: 161
   |||||||.....:.....:.....:.....:.....:.....:
377 CGGCTGTGTCTGCCCTTTCAGAGAGTCTTGCCCAAGTCGGGGGG 426
   |||||||.....:.....:.....:.....:.....:.....:
161 .....:.....:.....:.....:.....:.....: 161
   |||||||.....:.....:.....:.....:.....:.....:
427 CTGACCCAGAGCTTGCGTCCAGGACAAATCCCTACCGTG.....:.....: 467
   |||||||.....:.....:.....:.....:.....:.....:
162 .....:.....:.....:.....:.....:.....: 166
   |||||||.....:.....:.....:.....:.....:.....:
468 .....CTGAGTGGCGTGAACGTGCGGTGGTGTCTGAGGAG 504
   |||||||.....:.....:.....:.....:.....:.....:
166 nPheThrAspPheLeuTyrCysValAsnLeuLysLeuLeuProAsnGln 183
   |||||||.....:.....:.....:.....:.....:.....:
183 sPcysAlaLysAlaHisIleGluLysValThrAspAlaMetLeuLysAla 199
   |||||||.....:.....:.....:.....:.....:.....:
505 TCTGAGTAACTATGACCCGCTGTACACACCCAGCATGTTCTGCGCC 554
   |||||||.....:.....:.....:.....:.....:.....:
555 GCGGAGGCGCAAGACCAAGAGACTCCGCAACGCTGACTCTGGGGGGCC 604
   |||||||.....:.....:.....:.....:.....:.....:
200 GlyLysMetAspGlyGlyLysAspThrCysLysGlyAspSerGlyGlyPr 216

```

```

605 CCTGATCTGCACAGGGTACTTGCAGGGCCTTGCTCTTTCGAAAAAGCC 654
   |||||||.....:.....:.....:.....:.....:.....:
216 OLeuIleCysAspGlyValLeuGlnGlyIleThrSerTrpGlyHisThrP 233
   |||||||.....:.....:.....:.....:.....:.....:
655 CGTGTGGCCAAAGTTGGCGTCCAGGTGTCTACACCAACTCTCGCAATTTC 704
   |||||||.....:.....:.....:.....:.....:.....:
233 rOCysGlyGluProAspMetProGlyValTyrThrLysLeuAsnLysPhe 249
   |||||||.....:.....:.....:.....:.....:.....:
705 ACTGAGTGAATAGCAAAACCGTC 728
   |||||||.....:.....:.....:.....:.....:.....:
250 ThrSerTrpIleLysAspTrpMet 257
   |||||||.....:.....:.....:.....:.....:.....:
seq_name: p1r2:S55066

```

seq\_documentation\_block:

trypsin (EC 3.4.21.4) II precursor, pancreatic (clone 2-p29) - chicken

N:Alternate names: trypsinogen II

C:Species: Gallus gallus (chicken)

C:Date: 23-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 08-Sep-1997

C:Accession: S55066; S72347

R:Wang, K.; Gan, L.; Lee, L.; Hood, L.

Biochem. J. 307, 471-479, 1995

A:Title: Isolation and characterization of the chicken trypsinogen gene family.

A:Reference number: S55065

A:Accession: S55066

A:Molecule type: mRNA

A:Residues: 1-248 <WANI>

A:Cross-references: EMBL:U15157; NID:g603906; PID:g603907

A:Experimental source: Clone 2-P29

A:Accession: S72347

A:Molecule type: DNA

A:Residues: 1-248 <WANI>

A:Cross-references: EMBL:U15157; NID:g603906; PID:g603907

A:Experimental source: clone 2-P29

C:Suprafamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; protein digestion; serine proteinase; zymogen

F:1-16/Domain: signal sequence #status predicted <Sig>

F:17-25/Domain: activation peptide #status predicted <APT>

F:26-248/Product: trypsin II #status predicted <MAT>

F:26-241/Domain: trypsin homology <TRY>

F:65.109.202/Active site: His, Asp, Ser #status predicted

alignment\_scores:

Quality: 450.00 Length: 243

Ratio: 2.795 Gaps: 6

Percent Similarity: 66.255 Percent Identity: 38.683

alignment\_block:

US-09-030-606-173 x S55066 ..

Align seg 1/1 to: S55066 from: 1 to: 248

```

10 CACTGGCAGCCCTGGCAGGCGGCATGTCATGGAAGAAATTTGTTCTG 59
   |||||||.....:.....:.....:.....:.....:.....:
35 HisSerValProTyrGlnValSerLeu...AsnSerGlyTyrHisPheCy 50
   |||||||.....:.....:.....:.....:.....:.....:
60 CTGGGCGCTCTGTCATCCGACAGTGGGTGCTGCACCGCACACTGTT 109
   |||||||.....:.....:.....:.....:.....:.....:
50 sGlyGlySerLeuIleAsnSerGlnTrpValLeuSerAlaAlaHisCysT 67
   |||||||.....:.....:.....:.....:.....:.....:
110 TCACGACTCCTACACCATCGGGCTGGCGCTGCACAGTCTTGAGCCGAC 159
   |||||||.....:.....:.....:.....:.....:.....:
67 YrLysSerArgGlnIleGlnValArgLeuGlyGluTyrAsnIleAspVal... 82
   |||||||.....:.....:.....:.....:.....:.....:
160 CAAGAGCAGGAGGAGCAGATGCTGAGGCCAGCCCTCCGTAAGCCGACCC 209
   |||||||.....:.....:.....:.....:.....:.....:
83 GlnLysAspSerGluValAlaValArgSerSerSerValIleIleArgHisPr 99
   |||||||.....:.....:.....:.....:.....:.....:
210 AGATACACAGACCCCTGCTGCTGCTAACGACCTCATCTATCAAGTTGG 259
   |||||||.....:.....:.....:.....:.....:.....:
99 OlyTyrSerSerIleThrLeuAsnAsnAspIleMetLeuIleLysLeuA 116

```

```

260 ACGAATCCGTCGTCGAGTCGACACCATCCGAGCATCAGATTCCTTCG 309
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
116 laseAlaValGluTyrSerAlaAspIleGlnProIleAlaLeuProSer 132
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
310 CAGTCCCTACCCGGGAGACTTTCCTCTTCCTTCGCTGGGGT...CT 356
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
133 SerCysAlaValAlaGlyThrGluCysLeuIleSerGlyTyrGlyAsnThr 149
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
357 GCTGCGAAGCGGTGAGCTCAGCGGTGTGTGTGCTTCCTTCAGAGAGT 406
      |||:::|||||
149 rLeuSerAsnGly..... 153
407 CCTCTGCCAGTCGCGGGGGGTGACCCAGAGCTCTGCTCCAGGACAGA 456
      :::::
154 .....Tyr..AsnT 156
457 TCCCTACCGCTGCTGCAAGCTGTAACGTGTGCGGTGTCTGAGAGGTC 506
      |||  :::::|||||  |||  |||  |||  |||  |||  |||
156 YrProGluLeuLeuGlnCysLeuAsnAlaProIleLeuSerAspGln 172
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
507 TGCATAGCTCTATGACCCGCTGATACACCCACATCTTCGCGCGG 556
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
173 CysGlnGluAlaTyrProGlyAspIleThrSerAsnMetIleCysValG 189
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
557 CGGAGGCGACAGACAGAGAGACTCTGCAACGCTGCGGGGGGCCCC 606
      |  :::::|||||  |||  |||  |||  |||  |||  |||  |||
189 YrPheLeuGluGlyGlyLysAspSerCysGlnGlyAspSerGlyGlyPro 206
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
607 TGAATTCGACGCGGTACTTTCGAGGCGCTTGTCTTCGGAAGCCCCG 656
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
206 AlValCysAsnGlyGluLeuGlnGlyIleValSerTyrP...IleGly 221
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
657 TGTGGCCAAAGTGGCGTGCAGCTGTCTACCAACCTGCAATTCAC 706
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
222 CysAlaLeuLysGlyTyrProGlyValTyrThrLysValCysAsnTyrVa 238
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
707 TGAGTGTATAGAAAACCGTCCAGGCC 734
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
238 lAspTyrIleGlnGluThrIleAlaAla 247

```

```

seq_name: p1r2:A29586
seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) hGK-1 precursor - human
  N:Alternate names: glandular kallikrein
  C:Species: Homo sapiens (man)
  C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 20-Mar-1998
  C:Accession: A29586
  R:Schedlich, L.J.; Bennetts, B.H.; Morris, B.J.
  DNA 6, 429-437, 1987
  A:Title: Primary structure of a human glandular kallikrein gene.
  A:Reference number: A29586; MUID:88054467
  A:Accession: A29586
  A:Molecule type: DNA
  A:Residues: 1-261 <SCCH>
  A:Cross-references: GB:M18157; NID:9186640; PID:9386842
  A:Note: the authors translated the codon TAC for residue 43 as Trp
  C:Genetics:
  C:Intons: 16/1; 69/2; 165/1; 210/3
  C:Superfamily: trypsin; trypsin homology
  C:Keywords: hydrolase; serine proteinase
  F:25-253/Domain: trypsin homology <TRY>
  F:65,120,213/active site: His, Asp, Ser #status predicted

```

```

alignment_scores:
  quality: 449.50      length: 254
  ratio: 2.863         gaps: 5
  percent similarity: 61.811      percent identity: 37.402
alignment_block:
  US-09-030-606-173 x A29586

```

```

Align seg 1/1 to: A29586 from: 1 to: 261
10 CAGTCCGACCCCTGGCAGGCGGCACTGTCATGGAAGCAATTCCTTCG 59
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
34 HisSerGlnProTyrGlnValAlaValTyrSerHisGlyTyrAlaHisCys 50
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
60 CTCGGGCGCTCTGTCATCCGAGTGGGTGCTGTCACGCCACACTGT 109
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
50 SGLyGlyValLeuValHisProGlnTyrValLeuThrAlaHisCysL 67
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
110 TCCAGACTCTTACACATCGGGCTGGGCTGCACAGTCTTGAGGCCAG 159
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
67 euLysLysAsnSerGlnValTyrLeuGlyArgHisAsnLeuPheGluPro 83
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
160 CAGAGCCAGGAGCCAGATGTGTGAGGCCAGCCCTTCGTAAGCCACC 209
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
84 GluAspThrGly...GlnArgValProValSerHisSerPheProHisPr 99
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
210 AGAGTACAAACAGACCTTGCTC..... 231
      |  |||:::|||||  |||  |||  |||  |||  |||  |||  |||
99 OleuTyrAsnMetSerLeuLeuLysHisGlnSerLeuArgProAspGlu 116
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
232 .....GCTAAGACCTCATGCTCATCAAGTTGAGAGAAATCCGTCCAG 276
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
116 sPserSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLysIle 132
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
277 TGTGACACCATCCGAGCATCAGCATTCCTTCGACAGTCCCTACCGCGG 326
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
133 ThrAspValValLysValLeuGlyLeuProThrGlnGluProAlaLeuG 149
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
327 GAACCTTCGCGCTGTCGCTGGGTGGGTGCTGTCGGAAGGTAGTCA 376
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
149 YThrThrCysTyrAlaSerGlyTyrPheLysIle..Glu..... 161
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
377 CGGCTGTGTCTGCCCTTCACAGAGAGTCTTCGACAGTCCGCGGGGG 426
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
161 ..... 161
427 CTGACCCAGAGCTTCGCTCCAGGACAGAAATCCATCCGTGTCAGTGC 476
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
162 ....ProGluGluPheLeu.....ArgProArgSerLeuGlnCys 173
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
477 GTGAGGTGTGTGCTGTCGTCGAGAGGTCGACATAGCTCTATGACCC 526
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
174 ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTyrSerG 190
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
527 GCTGTACACCCACAGATTCCTGCGCGCGGAGGCGACAGCAGAAAG 576
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
190 uLysValThrGluPheMetLeuCysAlaGlyLeuThrPheThrGlyLysA 207
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
577 ACTCTGCAACGCTGTCGTGGGGGCGCCCTGATCTGCAACGGTACTTG 626
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
207 sPThrCysGlyGlyLysAspSerGlyProLeuValCysAsnGlyValLeu 223
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
627 CAGGCGCTTGTGCTTCGGAAACCCCGGTGGCCCAAGTTGGCGCTCC 676
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
224 GlnGlyIleThrSerTyrPheLysProGluProCysAlaLeuProGluLysPr 240
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
677 AGGTGCTACACCAACCTCTGCNAATTCAGTGGATGATAGAAACCG 726
      |||:::|||||  |||  |||  |||  |||  |||  |||  |||
240 oAlaValTyrThrLysValAlaHisTyrArgLysTyrPheLysAspThrI 257
      :::::|||||  |||  |||  |||  |||  |||  |||  |||
727 TCCAGGCCAGT 737
      :::::|||||
257 leAlaAlaAsn 260

```

```

seq_name: p1r2:A31136
seq_documentation_block:
  tissue kallikrein (EC 3.4.21.35) 7 precursor, submandibular - rat
  N:Alternate names: glandular prokallikrein 7, submandibular; proteinase A
  C:Species: Rattus norvegicus (Norway rat)
  C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 20-Mar-1998

```

C:Accession: A31136; S10698; S10699; D41429; B41429; S09315  
 R:Chen, Y.P.; Chao, J.; Chao, L.  
 Biochemistry 27, 7189-7196, 1988  
 A:Title: Molecular cloning and characterization of two rat renal kallikrein genes.  
 A:Reference number: A31136; MUID:89088074  
 A:Accession: A31136  
 A:Molecule type: DNA  
 A:Residues: 1-261 <CHE>  
 A:Cross-references: GB:J02837; NID:920499; PID:920500  
 R:Elmoujhed, A.; Gutman, N.; Brillard, M.; Gauthier, F.  
 FEBS Lett. 265, 137-140, 1990  
 A:Title: Substrate specificity of two kallikrein family gene products isolated from the  
 A:Accession: S10698; MUID:90306305  
 A:Accession: S10698  
 A:Molecule type: protein  
 A:Residues: 25-36 <ELM>  
 A:Accession: S10699  
 A:Molecule type: protein  
 A:Residues: 112-139 <EL2>  
 R:Kato, H.; Nakanishi, E.; Enjoji, K.; Hayashi, I.; Oh-Ishii, S.; Iwanaga, S.  
 J. Biochem. 102, 1389-1404, 1987  
 A:Title: Characterization of serine proteinases isolated from rat submaxillary gland: w  
 A:Reference number: A41429; MUID:88198057  
 A:Accession: D41429  
 A:Molecule type: protein  
 A:Residues: 112-133 <KAT>  
 A:Accession: B41429  
 A:Molecule type: protein  
 A:Residues: 25-34, D, 36-45, S', 47-67, X', 69-75 <KA2>  
 R:Brady, J.M.; MacDonald, R.J.  
 Arch. Biochem. Biophys. 278, 342-349, 1990  
 A:Title: The expression of two kallikrein gene family members in the rat kidney.  
 A:Reference number: S09315; MUID:90225801  
 A:Accession: S09315  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 43-45, S', 47-114, A', 116-261 <BRA>  
 C:Superfamily: trypsin; trypsin homology  
 C:Keywords: hydrolase; serine proteinase  
 F:1-18/Domain: signal sequence #status predicted <Sig>  
 F:19-261/Product: tissue prokallikrein 7, submandibular #status predicted <KAT>  
 F:25-253/Domain: trypsin homology <TRY>  
 F:65,120,213/Active site: His, Asp, Ser #status predicted

alignment\_scores:  
 Quality: 449.00 Length: 255  
 Ratio: 2.738 Gaps: 5  
 Percent Similarity: 64.314 Percent Identity: 36.863

alignment\_block:  
 US-09-030-606-173 x A31136 ..

Align seq 1/1 to: A31136 from: 1 to: 261

```

10 CACTCGAGCCCTGCGAGCGCGACATGATGAAACGATGTTCTG 59
   ::::::::::::::::::::::::::::::::::::
34 AsnserlnProrPglInValAlaLeuTyrSerPheThrIlySTYrLeuCY 50
   ::::::::::::::::::::::::::::::::::
60 CTGCGGGCTCTGTGTCATCCGAGTGGGTCTGTCAAGCCGACATGTT 109
   ::::::::::::::::::::::::::::::::::
50 SGlyGlyValLeuIleAspProSerTrpValIleThrAlaIleHisCys 67
   ::::::::::::::::::::::::::::::::::
110 TCAGACATCTCAACATCGGGCTGGCCGACAGCTTGAGGCGGAC 159
   ::::::::::::::::::::::::::::::::::
67 eRserAAsnTYrGlnValTrpLeuGlyArgAAsnAsnLeuLeuGlnAsp 83
   ::::::::::::::::::::::::::::::::::
160 CAAGAGCCAGGAGCCAGATGGTGGAGGCGAGCTTCGACGACGCC 209
   ::::::::::::::::::::::::::::::::::
84 ...GluRProPheAlaGlnHisArgLeuValSerInSerPheProHisP 99
   ::::::::::::::::::::::::::::::::::
210 AGACTACAGACAGCCCTGCTC..... 231

```

```

99 oAsPTyr...LysProPheLeuMetArgAsnHisThrArgLysProGlyA 115
232 .....GCTACGACCTATGCTCATCAAGTTGACGACGATCCGTGC 273
   ::::::::::::::::::::::::::::::::::
115 sPAsPHisSerAAsnAspLeuMetLeuIleHisLeuSerGlnProAlaAsp 131
   ::::::::::::::::::::::::::::::::::
274 GAGTGTGACACATCCGAGACATGACATGCTTCGACAGTCCCTACCG 323
   ::::::::::::::::::::::::::::::::::
132 IleThrAspGlyValIleAspLeuProThrGlnGluProLysVal 148
   ::::::::::::::::::::::::::::::::::
324 GGGGACTCTTGCCCTGTTCTGCTGGGTGCTGTCGGCGACGTTGAGC 373
   ::::::::::::::::::::::::::::::::::
148 lGlySerThrCysLeuAlaSerGlyTrpGly..... 158
374 TCACGGGTGTGTGTCGCCCTCTCAAGAGGTCCTGCCCCAGTGGCG 423
158 ..... 158
424 GGCGTCACCCAGAGCTCTGCGTCCAGGC.....AGATGCCCTACCGT 466
   ::::::::::::::::::::::::::::::::::
159 .....SerThrLysProLeuIleTrpGluPheProAspAs 170
467 GCTGACAGTGGGTAGAGTGTGGTGTGTGAGAGGTCTGACAGTAAGC 516
   ::::::::::::::::::::::::::::::::::
170 PleuGlnCysValAsnIleHisLeuSerAsnGlnLysCysIleLysH 187
517 TGTATGACCCGCTGTACACACCCAGATGTCTGCCCCGGCGAGGCA 566
   ::::::::::::::::::::::::::::::::::
187 lArTrLysGlnLysValThrAspLeuMetLeuLysAlaGlyLeuGln 203
567 GACCAGAGGAGCTCTGACAGGTGACTGGGGGCGCCCTGATGCA 616
   ::::::::::::::::::::::::::::::::::
204 GlYGlYrLysAspThrGlyAspSerGlyGlyProLeuLeuLysAs 220
617 CGGCTACTTCAGGGCCTTGTGTTGGAAAAAGCCCGTGGCCAG 666
   ::::::::::::::::::::::::::::::::::
220 pGlyValLeuGlnGlyIleThrSerTrpGlySerValProCysAlaLys 237
667 TTGGCGTCCAGAGTGTCTACCAACCTCTGCAATTCATGAGTGA 716
   ::::::::::::::::::::::::::::::::::
237 hRAsnMetProAlaIleTrpThrLysLeuIleLysPheThrSerTrp 253
717 GAGAAACCGTCCAG 731
   ::::::::::::::::::::::::::
254 LysGluValMetLys 258
seq_name: p1r2:JE0236
seq_documentation_block:
tissue kallikrein (EC 3.4.21.35) - mouse
N:Alternate names: prorenin-converting enzyme
C:Species: Mus musculus (house mouse)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 17-Mar-1999
C:Accession: JE0236
R:Hosoi, K.; Tada, J.; Tsumura, K.; Kanamori, N.; Yamana, N.
J. Biochem. 124, 368-376, 1998
A:Title: Expression of an allozyme of prorenin-converting enzyme in the submandibular
A:Reference number: JE0236; MUID:98351895
A:Accession: JE0236
A:Molecule type: mRNA
A:Residues: 1-261 <HOS>
A:Cross-references: DDBJ:AB016032
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
F:65,120,213/Active site: His, Asp, Ser #status predicted
F:102/Binding site: carbonylate (Asn) (covalent) #status predicted

```

alignment\_scores:  
 Quality: 448.50 Length: 252  
 Ratio: 2.786 Gaps: 6  
 Percent Similarity: 63.889 Percent Identity: 38.492

## alignment\_block:

US-09-030-606-173 x JE0236 ..

Align seg 1/1 to: JE0236 from: 1 to: 261

```

10 CACTCGACGCGCTGGAGCGGCGCACTGGTCATGGAAGCAATGTTCTG 59
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 AsnSerGlnProThrPheValAlaValIleValTyrGlnIleGlnIleCys 50
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 CTGGGCGCTCTGTCGATCCGAGTGGGTGCTGTCACAGCCGACATGTT 109
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 sGlyIleValLeuLeuAspArgAsnTrpValLeuThrAlaIleHisCys 67
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 TCAGAACTCTACACATCGGCGTGGGCTGACAGTCTTGAGGCGGAC 159
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 yValAspGlnTyrGlnValTrpLeuGlyLysAsnIleLeu...PheGln 82
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 CAAGAGCCAGGAGGAGGAGAGTGGAGGCGGAGGCGCTTCCGATGCGG 209
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 GluGlnProSerAlaGlnHisArgLeuValSerLysSerPheProHisPr 99
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 AGAGTACAAAGACCCCTGCTC..... 231
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
99 OGlyTyrAsnMetSerLeuLeuMetLeuGlnThrIleProProGlyAla 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
232 .....GCTAACGACCTCATGCTCATCAAGTTGAGCAATCCGTCGCGAG 276
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 sPheSerAsnAspLeuMetLeuLeuAlaGlyLeuSerLysProIleAsp 132
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
277 TCTGACACCATCCGAGAGCATCAGCATTCCTTCGACATCCGTCGCGG 326
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 ThrAspValValLysProIleAlaLeuProThrLysGlnProLysProG 149
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
327 GAACCTTGCTGCTGTTCTGGCTGGGCTGCTGCGGCAAGGTAAGTCA 376
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
149 ySerLysCysLeuValSerGlyTrpGlySerIle..... 160
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
377 CGGAGTGTGTGCTGCTCTTCAAGAGAGTCTGCGCCAGTCGCGGCGG 426
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 .....Thr.ProThrArgTyr.. 165
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
427 CTGACCCAGAGCTGTGCTCCAGGAGAGATGCCATCCGTCGTCAGTGC 476
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166 .....GlnLys.ProAspAspLeuGlnCys 173
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
477 GTGAACGTCGTGGTGTCTGAGAGAGTCTGCACTAAGCTCTATGACCC 526
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 ValPheIleThrLeuLeuProAsnGlnAsnCysAlaLysValTyrLeuG 190
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
527 GCTGACACACCCAGCATGTTCTGGCGGCGGAGGAGGCAAGCCAGAGG 576
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 nLysValThrAspValMetLeuCysAlaGlyLeuMetLysGlyLys 207
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
577 ACTCTCGACGAGTACTGTGGGGGCGCCCTGATCTGCAACGCGTACTTG 626
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 sPheThrCysArgAspAspSerGlyLysProLeuIleCysAspLysIleLeu 223
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
627 CAGGCGCTGTGTCTTGGAAAGCCCGGTGTGGCAAGTTGCGGTGCC 676
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 GlnGlyThrThrSerTyrGlyProThrProCysGlyLysProGlyValPr 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
677 AGGCTCTACACCACTCTGCAATTCATGATGAGTATAGAAACCG 726
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
240 oAlaIleThrThrAsnLeuIleLysPheAsnSerTrpIleLysAspThr 257
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
727 TC 728
   :
257 et 257

```

seq\_name: p1r2:S01971

seq\_documentation\_block:  
tissue kallikrein (EC 3.4.21.35) precursor - mouse

N:Alternate names: glandular kallikrein

C:Species: Mus musculus (house mouse)

C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 01-Aug-1997

C:Accession: S01971, I70023

R:Drinkwater, C.C.; Richards, R.I.

Nucleic Acids Res. 16, 10918, 1988

A:Title: Sequence of mGK-11, a mouse glandular kallikrein gene.

A:Reference number: S01971; M01D:89083511

A:Accession: S01971

A:Molecule type: DNA

A:Residues: 1-261 &lt;DR1&gt;

A:Cross-references: EMBL:X13215

R:Evans, B.A.; Drinkwater, C.C.; Richards, R.I.

J. Biol. Chem. 262, 8027-8034, 1987

A:Title: Mouse glandular kallikrein genes: structure and partial sequence analysis of

A:Reference number: I55260; M01D:8750386

A:Accession: I70023

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 70-122 &lt;RES&gt;

A:Cross-references: GB:M18610; NID:q198529; PID:q198537

C:Genetics:

A:introns: 16/1; 69/2; 165/1; 210/3

C:Superfamily: trypsin, trypsin homology

C:Keywords: hydrolase; serine protease

F:1-17/Domain: signal sequence #status predicted &lt;SIG&gt;

F:18-24/Domain: propeptide #status predicted &lt;PRO&gt;

F:25-261/Product: tissue kallikrein #status predicted &lt;MAT&gt;

F:25-253/Domain: trypsin homology &lt;TRY&gt;

F:65,120,213/Active site: His, Asp, Ser #status predicted

## alignment\_scores:

Quality:	448.50	Length:	251
Ratio:	2.786	Gaps:	5
Percent Similarity:	64.143	Percent Identity:	38.247

## alignment\_block:

US-09-030-606-173 x S01971 ..

Align seg 1/1 to: S01971 from: 1 to: 261

```

10 CACTCGACGCGCTGGAGCGGCGCACTGGTCATGGAAGCAATGTTCTG 59
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 AsnSerGlnProThrPheValAlaValTyrArgTyrAsnLysTyrIleCys 50
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
60 CTGGGCGCTCTGTCGATCCGAGTGGGTGCTGTCACAGCCGACATGTT 109
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 sGlyIleValLeuLeuAspArgAsnTrpValLeuThrAlaIleHisCys 67
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
110 TCAGAACTCTACACATCGGCGTGGGCTGACAGTCTTGAGGCGGAC 159
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 sValSerGlnTyrAsnValTrpLeuGlyLysThrLysLeu...PheGln 82
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
160 CAAGAGCCAGGAGGAGGAGTGGAGGCGGAGGCGCTTCCGATGCGGAC 209
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 ArgGlnProSerAlaGlnHisArgMetValSerLysSerPheProHisPr 99
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
210 AGAGTACAAAGACCCCTGCTC..... 231
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
99 OGlyTyrAsnMetSerLeuLeuMetLeuGlnThrIleHisAsnProGlnProLys 116
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 sPheSerAsnAspLeuMetLeuLeuAlaGlyLeuSerGlnProIleAsp 132
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
277 TCTGACACCATCCGAGAGCATCAGCATTCCTTCGACATCCGTCGCGG 326
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 ThrAspValValLysProIleAlaLeuProThrLysGlnProLysLeuG 149
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
327 GAACCTTGCTGCTGTTCTGGCTGGGCTGCTGCGGCAAGGTAAGTCA 376
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
149 ySerThrCysLeuValSerGlyTrpGlySerIleThr..... 161

```

```
377 CCGGTGTGTCTGCCCTCTCAAGAGATCTCTGCCAGTCGGGGG 426
      |||:: |||
162 .....Prothr..Lys..... 164
427 CTGACCCAGAGCTCTGCTGCCAGAGCAATGCCCTACCGTGTGCAATGC 476
      :: ||| |||||
165 .....PheGlnThrProAspAspLeuGlnCys 173
477 GTGAACGTGTGGGTGTGTGTGAGAGGTCTGCAGTAAGCTATATGACCC 526
      |||:::.....:||||||| ||| ::::
174 ValSerIleLysLeuLeuProAsnGluValCysValLysAsnHisAsnG1 190
527 GCTGTACCAACCCAGCATGTCTGCGCGCGCGAGGGCAAGACCAGAAG 576
      |||:::||||||| |||
190 nLysValThrAspValMetLeuCysAlaGlyGluMetGlyGlyLysA 207
577 ACTCTGCAACGGTGACTCTGGGGGGCCCTGATCTGCAACGGGTACTTG 626
      ||:||||:||||||| ||||| |||
207 sPthrcysLysGlyAspSerGlyGlyProLeuIleCysAspGlyValLeu 223
627 CAGGGCCTGTGTCTTCGAAAGCCCGCTGTGCCCAAGTTGGCGTCC 676
      ::|||:::||||| ||||| |||
224 HisGlyIleThrAlaTrpGlyProIleProCysGlyLysProAsnThrPr 240
677 AGGTGTCTACACCAACCTCTGCAATTCAGTGAAGATAGAGAAACCG 726
      ||||| ||||| ||| ||||| ||||| |||
240 ogLysValTyrThrLysLeuIleLysPheThrAsnTrpIleLysAspThrW 257
727 TC 728
      ::
257 et 257
```

OM of: US-09-030-606-173 to: SwissProt\_37.\* out\_format: pfs  
Date: Sep 25, 1999 4:45 PM

About: Results were produced by the Gencore software, version 4.5,  
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:  
-MODE=lfname+02p\_model -DEV=xlp  
-O=/gen2.1/USPRO-seq/us09030606/runat\_24091999\_171618\_29683/app-query.fasta.1  
-DB=SwissProt\_37 -OPMT=fastan -SUFFIX=esp -GAPOP=12.000  
-GAPEXT=0.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=0.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-DEL0P=0.000 -XGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR\_SCORE=score  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09030606  
-NCPU=6 -ICPU=3 -WAIT -THREADS=1

Search information block:  
Query: us-09-030-606-173  
Query length: 1265  
Database: SwissProt\_37.\*  
Database sequences: 77977  
Database length: 28268293  
Search time (sec): 155.430000

score\_list:  
Sequence Strd Orig Zscore Escore Len Documentation  
SwissProt\_37:SCCE\_HUMAN + 519.50 500.54 1.0e-20 253 | P49862 rat mus sapiens (human).  
SwissProt\_37:KLK4\_MOUSE + 459.00 443.92 1.5e-17 244 | P36376 rat mus norvegicus (rat).  
SwissProt\_37:TRRY\_CANFA + 457.00 441.99 1.9e-17 246 | P06871 canis familiaris (dog).  
SwissProt\_37:KLK4\_MOUSE + 455.00 439.75 2.4e-17 259 | P36376 rat mus norvegicus (rat).  
SwissProt\_37:KLK4\_MOUSE + 452.50 437.34 3.2e-17 261 | P36376 rat mus norvegicus (rat).  
SwissProt\_37:TRRY\_MOUSE + 452.00 437.26 3.4e-17 247 | P08426 rat mus norvegicus (rat).  
SwissProt\_37:KLK4\_MOUSE + 451.50 436.46 3.6e-17 259 | P15948 mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 450.00 434.99 4.3e-17 261 | P00756 mus musculus (mouse).  
SwissProt\_37:TRRY\_CHICK + 450.00 435.35 4.4e-17 248 | P09029 gallus gallus (chick).  
SwissProt\_37:KLK4\_MOUSE + 449.50 434.52 4.6e-17 261 | P36376 rat mus norvegicus (rat).  
SwissProt\_37:KLK4\_MOUSE + 449.00 434.05 4.9e-17 261 | P15946 mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 448.50 433.58 5.2e-17 261 | P15946 mus musculus (mouse).  
SwissProt\_37:PROS\_HUMAN + 448.50 433.58 5.2e-17 261 | P07288 homo sapiens (human).  
SwissProt\_37:TRRY\_MOUSE + 448.00 433.53 5.6e-17 246 | P32822 rat mus norvegicus (rat).  
SwissProt\_37:PROS\_HUMAN + 448.00 433.53 5.6e-17 244 | P09287 homo sapiens (human).  
SwissProt\_37:KLK4\_MOUSE + 446.50 431.70 6.1e-17 261 | P15949 mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 446.50 429.82 8.4e-17 261 | P36376 rat mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 443.00 428.52 1.0e-16 257 | P07276 macaca fascicularis (macaca).  
SwissProt\_37:TRRY\_CHICK + 443.00 428.77 1.0e-16 248 | P09029 gallus gallus (chick).  
SwissProt\_37:TRRY\_CHICK + 441.00 426.89 1.3e-16 248 | P09029 gallus gallus (chick).  
SwissProt\_37:TRRY\_MOUSE + 441.00 426.95 1.3e-16 246 | P32822 rat mus norvegicus (rat).  
SwissProt\_37:TRRY\_XENLA + 441.00 426.95 1.3e-16 244 | P70059 xenopus laevis (afrod).  
SwissProt\_37:TRRY\_MOUSE + 440.50 426.45 1.4e-16 247 | P12788 rat mus norvegicus (rat).  
SwissProt\_37:TRRY\_MOUSE + 439.50 425.12 1.5e-16 261 | P00758 rat mus norvegicus (rat).  
SwissProt\_37:PROS\_HUMAN + 439.50 425.12 1.5e-16 261 | P33619 macaca mulatta (rhesus).  
SwissProt\_37:KLK4\_MOUSE + 439.50 425.12 1.5e-16 259 | P00758 rat mus norvegicus (rat).  
SwissProt\_37:TRRY\_MOUSE + 438.00 424.21 1.9e-16 243 | P00760 bos taurus (bovine).  
SwissProt\_37:KLK4\_MOUSE + 437.50 423.24 2.0e-16 261 | P36376 rat mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 437.50 423.32 2.0e-16 258 | P00756 mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 435.00 420.89 2.7e-16 261 | P15947 mus musculus (mouse).  
SwissProt\_37:TRRY\_HUMAN + 432.00 418.46 3.8e-16 247 | P07477 homo sapiens (human).  
SwissProt\_37:TRRY\_MOUSE + 431.00 417.99 4.4e-16 231 | P00761 sus scrofa (pig).  
SwissProt\_37:TRRY\_MOUSE + 430.00 416.61 4.9e-16 246 | P00762 rat mus norvegicus (rat).  
SwissProt\_37:TRRY\_MOUSE + 429.00 415.25 5.5e-16 261 | P15945 mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 428.00 414.31 6.2e-16 261 | P00755 mus musculus (mouse).  
SwissProt\_37:KLK4\_MOUSE + 426.00 412.41 7.8e-16 262 | P06870 homo sapiens (human).  
SwissProt\_37:TRRY\_MOUSE + 424.00 410.97 1.0e-15 246 | P07146 mus musculus (mouse).  
SwissProt\_37:TRRY\_MOUSE + 423.50 410.94 1.1e-15 231 | P35032 salmo salar (atlantic).  
SwissProt\_37:TRRY\_MOUSE + 422.00 409.32 1.3e-15 238 | P35033 salmo salar (atlantic).  
SwissProt\_37:TRRY\_MOUSE + 420.00 407.18 1.6e-15 243 | P09463 bos taurus (bovine).  
SwissProt\_37:TRRY\_MOUSE + 419.50 406.82 1.7e-15 247 | P19799 xenopus laevis (afrod).  
SwissProt\_37:TRRY\_MOUSE + 419.50 407.24 1.7e-15 229 | P00764 salmo salar (atlantic).

seq\_name: SwissProt\_37:SCCE\_HUMAN  
seq\_documentation\_block:  
ID SCCE\_HUMAN STANDARD; PRT; 253 AA.  
AC P49862;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME PRECURSOR (EC 3.4.21.-) (SCCE).  
GN PRSS6 OR SCCE.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.  
RC TISSUE=SKIN.  
RX MEDLINE: 94308225.  
RA HANSSON L., STROEMQVIST M., BAECKMAN A., WALLBRANDT P., CARLSTEIN A.,  
RA EGELRUD T.;  
RT "Cloning, expression, and characterization of stratum corneum  
RT chymotryptic enzyme. A skin-specific human serine proteinase.";  
RL J. BIOL. CHEM. 269:19420-19426(1994).  
RN [2]  
RX CHARACTERIZATION.  
RX MEDLINE: 95314630.  
RA SKYTT A., STROEMQVIST M., EGELRUD T.;  
RT "Primary substrate specificity of recombinant human stratum corneum  
RT chymotryptic enzyme.";  
RN [3]  
RX BIOTHER. BIOPHYS. RES. COMMUN. 211:586-589(1995).  
RT "- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE  
RT STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS  
RT SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID  
RT RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE  
RT CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-  
RT 1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION  
RT OF PRECURSORS TO INFLAMMATORY CYTOKINES.  
RT "- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS  
RT EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE  
RT ALSO SEEN IN THE BRAIN AND KIDNEY.  
RT "- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
RT TRYPSIN FAMILY.  
-----  
CC THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION  
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -  
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS  
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL  
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL: L33404; G532504; -  
DR PROSITE: PS00134; TRYPsin\_HIS; 1.  
DR PROSITE: PS00135; TRYPsin\_SER; 1.  
DR PFAM: PF00089; trypsin; 1.  
KW HSDOLASE; SERINE PROTEASE; ZYMOGEN; GLYCOPROTEIN; SIGNAL.  
FT SIGNAL 1 22  
FT PROPEP 23 22  
FT CHAIN 30 253  
FT ACT\_SITE 70 70  
FT ACT\_SITE 112 112  
FT ACT\_SITE 205 205  
FT DISULFID 36 137  
FT DISULFID 55 71  
FT DISULFID 144 211  
FT DISULFID 176 190  
FT DISULFID 201 226  
FT CARBOHYD 246 246  
FT SEQUENCE 253 AA; 27525 MW; 07FDB9F7 CRC32;

alignment\_scores:                   Quality: 519.50                   Length: 247  
                                   Ratio: 3.038                   Gaps: 5  
 Percent Similarity: 69.231           Percent Identity: 42.915

alignment\_block:  
 US-09-030-606-173 x SCCE\_HUMAN                   ..

Align seg 1/1 to: SCCE\_HUMAN from: 1 to: 253

```

1 GGCAGCCG.....CACTGCAGCCCTGGAGCGGCGACTGTGAT 41
   |||||
33 GlyAlaProCysAlaIaArgGlySerHisProTrpGlnValAlaLeuLeu 49
   |||||
42 GGAAGAAGATGTTGCTGCTGGCGCTCTGTGATCCGAGTGGTGC 91
   |||||
49 rGlyAsnGlnLeuHisCysGlyGlyValLeuValAsnGlnGlyTrpVal 66
   |||||
92 TGTGAGCGGCACTGTTTCCAGAACTCTACACCATCGGCGGCTG 141
   |||||
66 eunrAlaAlaHisCysLysMetAsnGlnIuTrpValHisLeuGlySer 82
   |||||
142 CACAGTCTTGAGCGCCAGCAAGACCCAGGAGCCAGATGTTGAGCCAG 191
   |||||
83 AsphTrLeu...GlyAspArgArg.....AlaGlnArgIleLysAla 96
   |||||
192 CCTTCGCTAGCGGACCCAGAGTACAAACAGCCTTGCTGCTACGACC 241
   |||||
96 rLysSerPheArgHisProGlyTrpSerThrGlnThrHisValAsnAsp 113
   |||||
242 TCAAGTCTATCAATGTGAGCAATCCGTCGTCGATCGATGACCATCGG 291
   |||||
113 eunMetLeuValLysLeuAsnSerGlnAlaArgLeuSerMetValLys 129
   |||||
292 AGCATCAGCATTCCTTCGACAGTCCCTACCCGCGGGAACCTTTCCTCGT 341
   |||||
130 LysValArgLeuProSerArgCysGlnProProGlyThrTrpCysThrVa 146
   |||||
342 TTCGCTGCGGCTGCTGCTGCGAAGCGGTGACCTACGCGGTGTGTCTGC 391
   |||||
146 lSerGlyTrpLys..... 150
   |||||
392 CCTCTCAAGAGAGTCTCTGCGCCAGTCCGCGGCGGTGACCCAGAGCT 441
   |||||
151 ..... 154
   |||||
442 GCGTCCAGGAGAGATGCTACCGTCTGCTGAGTCCGTGAGCTGCGGTG 491
   |||||
155 ProAspAlaThr..PheProSerAspLeuMetCysValAspValLysLeu 170
   |||||
492 GTGCTGAGAGGTCTGCTGACATGACTATGACCCGCTGTACCAACCCAG 541
   |||||
171 lLeuSerProGlnAspCysThrLysValTrpLysAspLeuGlnAsnSe 187
   |||||
542 CATTTTTCGCGCGCGGAGGAGGAGCAAGAGCATCTGCAAGCGGTG 591
   |||||
187 rMetLeuCysAlaIleGlyLeuProAspSerLysLysAsnAlaCysAsnLys 204
   |||||
592 ACTCTGGGGGCGCCCTGATCTGCAACAGGATCTGACAGGCGCTGTGCT 641
   |||||
204 sPserGlyGlyProLeuValCysArgLysThrLeuGlnGlyLeuValSer 220
   |||||
642 TTCGAAAAGCCCGCTGTCGCAAGTGGCTGCGCAGGTGCTGTACACCAA 691
   |||||
221 TrpAlaThrPheProCysGlyGlnProAsnAspProGlyValIlyTrpG 237
   |||||
692 CCTCTGCAAAATTCAGTGTAGATAGAGAAAAGCGTCCAG 731
   |||||
237 nValCysLysPheThrLysTrpIleAsnAspThrMetLys 250

```

seq\_name: SwAspProc\_37:KLKA\_RAT

```

seq_documentation_block:
ID      KLKA_RAT      STANDARD:      PRT:      244 AA.
AC      P36375.
DT      01-JUN-1994 (REL. 29, CREATED)
DT      01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT      01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE      GLANDULAR KALLIKREIN 10 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE      (T-KININOGENASE) (K10) (PROTEINASE B) (ENDOPEPTIDASE K) (FRAGMENT).
CN      K10 OR K10-10.
OS      RATTUS NORVEGICUS (RAT).
OC      EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC      RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
NC      [1]
RP      SEQUENCE FROM N. A., AND PARTIAL SEQUENCE.
RC      TISSUE-SUBMANDIBULAR GLAND, AND KIDNEY;
RX      MEDLINE: 93041794.
RA      MA J.-X., CHAO J., CHAO L.;
RT      "Molecular cloning and characterization of rK10, a cDNA encoding T-
RT      kininogenase from rat submandibular gland and kidney.";
RL      BIOCHEMISTRY 31:10922-10928(1992).
RN      [2]
RP      SEQUENCE OF 10-32; 95-124 AND 179-232.
RX      MEDLINE: 91224135.
RA      GUTMAN N., ELMOUTARHD A., BRILLARD M., DU SORBIER B., GAUTHIER F.;
RT      "Microheterogeneity of rat submaxillary gland kallikrein k10, a
RT      member of the kallikrein family.";
RL      EUR. J. BIOCHEM. 197:425-429(1991).
RN      [3]
RP      SEQUENCE OF 10-32 AND 97-133.
RC      TISSUE-SUBMAXILLARY GLAND;
RX      MEDLINE: 90153911.
RA      XIONG W., CHEN L.-M., CHAO J.;
RT      "Purification and characterization of a kallikrein-like
RT      T-kininogenase.";
RL      J. BIOL. CHEM. 265:2822-2827(1990).
RN      [4]
RP      SEQUENCE OF 10-32 AND 97-117.
RX      MEDLINE: 88198057.
RA      KATO H., NAKANISHI E., ENJOJI K., HAYASHI I., OH-ISHI S., IWANAGA S.;
RT      "Characterization of serine proteinases isolated from rat
RT      submaxillary gland: with special reference to the degradation of rat
RT      kininogens by these enzymes.";
RL      J. BIOCHEM. 102:1389-1404(1987).
RN      [5]
RP      FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC      IN KININOGEN TO RELEASE LYS-BRADYKININ. THIS PROTEIN MAY BE
CC      INVOLVED IN THE REGULATION OF RENAL FUNCTION.
CC      -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC      SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC      KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC      MET-1-XAA OR LEU-1-XAA.
CC      -1- SUBUNIT: HETERO-DIMER OF A LIGHT CHAIN AND HEAVY CHAIN LINKED
CC      BY A DISULFIDE BOND.
CC      -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND, WHERE IT IS
CC      FOUND IN THE GRANULAR CONVOLUTED TUBULE AND STRIATED DUCT CELLS.
CC      IT IS LIKELY THAT THE ENZYME IS MAINLY SYNTHESIZED IN THE GRANULAR
CC      CONVOLUTED TUBULES AND THEN TRANSFERRED TO OTHER TISSUES BY
CC      RELEASE INTO THE VASCULATURE OR INTERSTITIAL SPACE.
CC      -1- PFM: PROBABLY N- AND O-GLYCOSYLATED. IT HAS CARBOHYDRATE MOETIES
CC      OF ALPHA-METHYL-D-MANNOSIDE AND N-ACETYL-D-GLUCOSAMINE GROUPS.
CC      -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC      TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: S48142; G259431; -.
CC      PIR: A35545; A35545.

```



DR PIR: B35545; B35545.  
 DR PIR: A44284; A44284.  
 DR PROSITE: PS00134; TRYPSIN\_HIS. 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER. 1.  
 DR PFAM: PF00089; trypsin. 1.  
 DR HSP: P00759; trypsin. 1.  
 DR HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN; SIGNAL;  
 KM MULTIGENE FAMILY.  
 FT SIGNAL 1 3 PROBABLE.  
 FT NON TER 1 3  
 FT PROPEP 4 9 ACTIVATION PEPTIDE (PROBABLE).  
 FT CHAIN 10 244 GLANDULAR KALLIKREIN 10.  
 FT CHAIN 10 96 T-KININOGENASE LIGHT CHAIN.  
 FT CHAIN 97 244 T-KININOGENASE HEAVY CHAIN.  
 FT ACT\_SITE 48 48 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 103 103 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 196 196 CHARGE RELAY SYSTEM.  
 FT ACT\_SITE 196 196 CHARGE RELAY SYSTEM.  
 FT DISULFID 16 156 BY SIMILARITY.  
 FT DISULFID 33 49 BY SIMILARITY.  
 FT DISULFID 135 202 BY SIMILARITY.  
 FT DISULFID 167 181 BY SIMILARITY.  
 FT DISULFID 192 217 BY SIMILARITY.  
 FT CARBOHYD 76 76 POTENTIAL.  
 FT CARBOHYD 91 91 POTENTIAL.  
 FT CARBOHYD 126 126 POTENTIAL.  
 FT CARBOHYD 142 142 POTENTIAL.  
 FT CONFLICT 29 29 N -> IET (IN REF. 3).  
 FT CONFLICT 115 116 IT -> DS (IN REF. 4).  
 FT CONFLICT 128 128 E -> G (IN REF. 3).  
 FT CONFLICT 133 133 S -> G (IN REF. 3).  
 SO SEQUENCE 244 AA; 27305 MW; 7661FB5F CRC32;

alignment\_scores:  
 Quality: 459.00 Length: 253  
 Ratio: 2.765 Gaps: 6  
 Percent Similarity: 65.613 Percent Identity: 39.130

alignment\_block:

US-09-030-606-173 x KLKA\_RAT ..

Align seg 1/1 to: KLKA\_RAT from: 1 to: 244

10 CACTCGACGAGCCCTGAGCGAGCGGCTGATGATGAAATGATGTCG 59  
 :: ||||| ::::|  
 19 AsnSerGlnProTyrPheValAlaIleIle.....AsnGluTyrLeu 33  
 60 CTGCGGCGCTGCTGTCATCCGAGTGGGCTGCTGACGCCGACACTGT 109  
 :: ||||| ::::|  
 33 sGlyGlyValLeuIleAspProSerTyrValIleThrAlaAlaHisCys 50  
 110 TCCCAACTCTTACACATCGGCGCTGAGCGGCTGACAGTCTTGAAGCCGAC 159  
 :: ||||| ::::|  
 50 ySerAsnTyrTrpHisValLeuLeuGlyArgAsnAsnLeuPheGluAsp 66  
 160 CAAGAGCAGGAGGAGCAGATGCTGAGGAGCGGCTCTCCGTCAGGACCC 209  
 :: ||||| ::::|  
 67 ..GluProPheAlaGlnTyrArgPheValAsnGlnSerPheProHisP 82  
 210 AGAGTACACAGACCCCTGCTC..... 231  
 :: ||||| ::::|  
 82 AspTyr...LysProPheLeuMetArgAsnHisThrArgGlnArgGly 98  
 232 .....GCTACAGACCTCATGCTCATCAATTTGAGCAGATCCGTGCC 273  
 :: ||||| ::::|  
 98 sPAspTyrSerAsnAspLeuMetLeuHisLeuSerGluProAlaAsp 114  
 274 GAGTGTGACACCATCGGAGCATGCTGCTGAGCAGTCCGTCACGCC 323  
 :: ||||| ::::|  
 115 IleHisAspGlyValIleValIleAspLeuProHisGluGluProLys 131  
 334 GGGAGACTTTCCTGCTTCTGCTGAGGCTGCTGTCGCGAAGCGTGAGC 373  
 :: ||||| ::::|

131 IGlySerThrCysLeuAlaSerGlyTyrGly..... 141  
 374 TCAGCGGTGTGTCTGCTGCCCTTCACAGAGATGCTCTGCCAGTCGCGG 423  
 142 .....SerThrLys 144  
 424 GGGGTGACCCA.GAGCTCTGCGTCCAGGAGCAAGATGCTACCGTGTGCA 472  
 :: ||||| ::::|  
 145 ProLeuAsnTyrGluLeu.....ProAspAspLeuGlu 155  
 473 GTGCGTGAACCTGTGCGGTGTGTCGAGAGGTCTGCAAGTCTATG 522  
 :: ||||| ::::|  
 155 ncysValAsnIleHisLeuSerAsnGluLysCysIleGluAlaTyrG 172  
 523 ACCGCGCTACACCCCGACATGTTCTGCGGCGGCGGAGGCGACAGACG 572  
 :: ||||| ::::|  
 172 IuGlnLysValThrAspLeuMetCysAlaGlyGluMetAspGlyArg 188  
 573 AAGCACTCTCAACAGGTGACTGTGGGCGCCCTGATCTCAACAGGTA 622  
 :: ||||| ::::|  
 189 LysAspThrCysLysGlyAspSerGlyProLeuIleCysAspGly 205  
 623 CTTCAGGCGCTGTGTCTTTCGGAAGCCCGTGTGCGCAAGTTGGCG 672  
 :: ||||| ::::|  
 205 IleuGlnIleThrSerTyrGlyAsnValProCysAlaGluProTyr 222  
 673 TGCCAGGTGTACACCACTCTGCAATTCATGATGATAGAGAA 722  
 :: ||||| ::::|  
 222 snProGlyValTyrThrLysLeuIleLysPheThrSerTyrIleLysGlu 238  
 723 ACCGTCACAG 731  
 :: ||||| ::::|  
 239 ValMetLys 241  
 :: ||||| ::::|  
 seq\_name: SwissProt\_37:TRY1\_CANFA  
 seq\_documentation\_block:  
 ID TRY1\_CANFA STANDARD; PRT; 246 AA.  
 AC P06871;  
 DT 01-JAN-1988 (REL. 06, CREATED)  
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
 DE TRYPSINOGEN, CATIONIC PRECURSOR (EC 3.4.21.4).  
 OS CANIS FAMILIARIS (DOG).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC CARNIVORA; FISSIPEDIA; CANIDAE; CANIS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 86284628.  
 RA PINSKY S.D., LAFORGE K.S., SCHEELE G.;  
 RT "Differential regulation of trypsinogen mRNA translation: full-length  
 RT mRNA sequences encoding two oppositely charged trypsinogen isoenzymes  
 RT in the dog pancreas."  
 RL MOL. CELL. BIOL. 5:2669-2676(1985).  
 CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPSIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M1590; G164097; -  
 DR PIR: B26273; TRDC.  
 DR PROSITE: PS00134; TRYPSIN\_HIS. 1.  
 DR PROSITE: PS00135; TRYPSIN\_SER. 1.  
 DR PFAM: PF00089; trypsin. 1.  
 DR HSP: P00761; 1EPT.

KW	HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZMOGEN; SIGNAL;			
HM	MULTIGENE FAMILY.			
FT	SIGNAL	1	15	ACTIVATION PEPTIDE.
FT	PROPEP	16	23	TRYPsin, CATIONIC.
FT	CHAIN	24	246	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	63	63	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	107	107	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	ACT_SITE	200	200	CHARGE RELAY SYSTEM (BY SIMILARITY).
FT	DISULFID	30	160	BY SIMILARITY.
FT	DISULFID	48	64	BY SIMILARITY.
FT	DISULFID	132	233	BY SIMILARITY.
FT	DISULFID	139	206	BY SIMILARITY.
FT	DISULFID	171	185	BY SIMILARITY.
FT	DISULFID	196	220	BY SIMILARITY.
FT	SITE	194	194	REQUIRED FOR SPECIFICITY (BY SIMILARITY)
SO	SEQUENCE	246 AA;	26170 MW;	DD106FSD CRC32;

  

alignment_scores:				
Quality:	457.00	Length:	245	
Ratio:	2.737	Gaps:	6	
Percent Similarity:	68.163	Percent Identity:	38.776	

  

alignment\_block:  
US-09-030-606-173 x TRY1\_CANFA ..

  

Align seg 1/1 to: TRY1\_CANFA from: 1 to: 246

4	ACCCGCGACTCGACCCCTGGCGAGCGCGACCTGGTCATGTGAAGAAGATT	53
31	SERATGASNSERVALPRDTGLINVALSERLEU...ASNSERGLYTRYH1	46
54	GTTTGCTCGGGCGTCTGGTGCATCCGACATGGGTGCTGCACCGCGAC	103
46	SPHECYSGLYGYSERLEU1EASNSERGLNTRYPA1VALSER1A1AH	63
104	ACTGTTTCCAGACTCTACACACATCGGGCGTGGCGCTGCACAGCTTGAG	153
63	ISCTYSTRYSERARGL1EGLNVALP1AGLEUGLYG1UTRYAS1H1E...	78
154	GCGGACCAAGAGCGAGGAGCGACAGTGTGTGAGGCCAGCCCTCGTAGC	203
79	AL1VALSERGLUGLYGLINPHE1EASNSER1A1A1ALYS1E1E1AR	95
204	GCACCCAGATGACACAGACCCCTTGCTGCTAACGACCTCATGCTCATCA	253
95	GL1SPROATGTRY1R1A1NAL1A1A1NTH1R1LE1ASP1A1A1SPL1EW1ET1E1UL	112
254	ACTTGAGCAATCCGCTGTCGCGAGTGTGACACATCCGGAGATAGACATT	303
112	Y1E1US1E1SER1PRO1AL1E1HR1E1U1ASNS1ER1AR1VAL1SER1A1L1E1A1L1E1U	128
304	GCTTGACAGTCCCTACGCGGGGGAACCTTGCTGCTTGTGCTGGGG	353
129	PRO1Y1S1SER1CY1SPRO1AL1A1E1GLY1HR1D1N1YS1LEU1E1SER1GLY1TP1E1L	145
354	TCTGCTGCGCAAGCGGTGAGCTCAGCGGTGTGTCTGCTGCTTCAAGGA	403
145	Y.....	145
404	GCTCTGTCGCGACATCGCGGGGCGTGCACCCAGAGACTGTGCTGCCAGCA	453
146	.....AS1NTH1R1N1E1R1E1GLY.....GL1N.....A	153
454	GANTGCGTACACGCTGCTGCGAGTGGCGAAGCTGTGGGTGTGAGGAG	503
153	STR1PR1O1A1SP1VAL1EUG1N1C1YS1LEU1Y1S1A1A1P1ROL1E1L1E1SER1A1SP1E1R	169
504	GTTTCCAGTACATCTATAGCCCGCTGTACCAACCCAGCATGTTCTCGC	553
170	VALCY1S1AR1GAS1NAL1AT1Y1R1PRO1GL1N1E1SER1E1ASNS1MET1E1CY1S1E	186
554	CGGGGAGGCGCAAGCACAGAAAGACCTCTGCACAGCGTGATCTTGCGGGGC	603

```

186 uqlylrmctgluglyglylshspsercysgngnglyaspseglglylp 203
|||||
203 rovalValLysasnglygluleuglnclyValValserirpely..Ala 218
|||||
654 CCCTGTGTCGCAACGGGCTATTCAGAGCGCTGTCTTCCGAAAGCC 653
|||||
203 rovalValLysasnglygluleuglnclyValValserirpely..Ala 218
|||||
219 GLYCYGALAGlnLysglYlpsProglYValserProglYValcyslYsty 235
|||||
704 CACTGAGTGATAGAGAAACGTCACGCGCAGT 737
|||||
235 rValserirpIleGlInhrIleAlaIasn 246
|||||

seq_name: SwissProt_37:KLKB_RAT

seq_documentation_block:
ID KLKB_RAT STANDARD: PRT: 255 AA.
AC P36376;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 12, SUBMANDIBULAR/RENAL PRECURSOR (EC 3.4.21.35)
DE (TISSUE KALLIKREIN) (RSG-3).
GN KLK12 OR KLK-12.
OS RATTUS NORVEGICUS (RAT).
OC EUARCHOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE: 89088074.
RA CHEN Y.-P., CHAO J., CHAO L.;
RT "Molecular cloning and characterization of two rat renal kallikrein
RL genes."
RT BIOCHEMISTRY 27:7189-7196(1988).
CC -I- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRAADYKININ.
CC -I- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LXSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-1-XAA OR LEU-1-XAA.
CC -I- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation --
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and that statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M19648; G205004; -
DR EMBL: M29222; G205004; JOINED.
DR PIR: B31136; B31136.
DR PROSITE: PS00134; TRYPsin_HIS.1.
DR PROSITE: PS00135; TRYPsin_SER.1.
DR PFM: PFM0089; trypsin.1.
DR HSP: P00759; ITON.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZMOGEN; SIGNAL.
FT SIGNAL 1 18
FT PROPEP 19 24 PROBABLE.
FT CHAIN 25 259 GLANDULAR KALLIKREIN 12.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
FT ACT_SITE 211 211 CHARGE RELAY SYSTEM.
FT DISULFD 31 171 BY SIMILARITY.
FT DISULFD 48 64 BY SIMILARITY.
FT DISULFD 150 217 BY SIMILARITY.
FT DISULFD 182 196 BY SIMILARITY.

```



Percent Similarity: 63.095 Percent Identity: 38.492

## Alignment block:

US-09-030-606-173 x KLR8\_RAT

Align seg 1/1 to: KLR8\_RAT from: 1 to: 261

```

10 CACTCGCAGCCCTGGCAGGCGGACATGTCATGGAAGAAAGAAATGTTCTG 59
   ::::::::::::::::::::::::::::
34 AsnserginProtrpGlnValAlaValIytrHisphasnngluProglncy 50
   ::::::::::::::::::::::::::::::
60 CTGCGGCGTCTGTCATCGCAGTGGGTCTGTACAGCCGACACTGTT 109
   ::::::::::::::::::::::::::::::
50 sgllyglValleuIleHisProserItrpValIleThrAlaIleHisCysT 67
   ::::::::::::::::::::::::::::::
110 TCAGAACTCTTACACCATGGGCGCCGACAGTCTGAGGCGGAC 159
   ::::::::::::::::::::
67 YrserValAsnIytrGlnValItrpLeuGlyArGAsnAsnLeuLeuGluAsp 83
   ::::::::::::::::::::::::::::
160 CAGAGCCAGGAGCCAGATGTCGAGGCGGACCTCTCCGTCAGGACCC 209
   ::::::::::::::::::::
84 ..GluProphleuIleHisArgLeuValSerGlnSerIleProHisSpr 99
   ::::::::::::::::::::
210 AGACTACACAGACCCTGCTC..... 231
   ::::::::::::::::::::
99 OGlyPhasnLeuAspIleIleIytrAsnHisThrArgIySprGlyAsnA 116
   ::::::::::::::::::::
232 ....GCTAACGACCTCATGCTCATCAATGAGAGAAATCGGTGCGGAG 276
   ::::::::::::::::::::::::::::
116 sptYrserAsnAspLeuLeuLeuHisLeuIytrHisProAlaAspIle 132
   ::::::::::::::::::::::::::::
277 TCTGACACATCCGAGCATCGACATGCTCTCGCATGTCCTTACGCGG 326
   ::::::::::::::::::::
133 ThrAspGlyValIyValIleAspLeuProthrGluGluProIyValGlu 149
   ::::::::::::::::::::
327 GAACCTTTCCTGCTTCTGCTGGGTCTGTGTCGCGAGACGTCAGCTCA 376
   ::::::::::::::::::::
149 YserThrCysLeuThrSerIytrpGlySerIleThr..... 161
   ::::::::::::::::::::
377 CGGCTGTGTCTGCTGCTCTCAAGAGAGTCCTTGCCTCAGTCCGCGGG 426
   ::::::::::::::::::::
162 .....Pro..LeuIyS..... 164
   ::::::::::::::::::::
427 CTGACCCAGAGCTGTGCTGCCAGAGAAATGCTTACCTGCTGTCAGTGC 476
   ::::::::::::::::::::
165 .....TrrpIuPherProAspAspLeuGlnCys 173
   ::::::::::::::::::::
477 GTGACAGTGTGCTGTGTCGAGAGGTCTGCAAGTAAGCTATGACCC 526
   ::::::::::::::::::::
174 ValAsnIleHisLeuLeuSerAsnGluIyScyIleIySalatIyAsnAs 190
   ::::::::::::::::::::
527 GCTGTACACCCAGCATGTCGCGCGGCGGAGGAGCAACAGCAAGG 576
   ::::::::::::::::::::
190 pgluValItrHispValMetLeuCysAlaGlyGluMetAspIyIyIySa 207
   ::::::::::::::::::::
577 ACTCTGCAAGGTCGTCGAGGCGGCGCTGATCTGCAAGGTCGTCAGT 626
   ::::::::::::::::::::
207 sptIecysIySglYAspserIyglYProleuIleCysAspGlyValIleu 223
   ::::::::::::::::::::
627 CAGGCGCTTGTCTTTCGAAAGCCCGTGTGCGCAAGTGGCGGTGCC 676
   ::::::::::::::::::::
224 GlnIyIleThrSerItrpGlySerMetItrpCysGlyIuProAsnIySpr 240
   ::::::::::::::::::::
677 AGGTCGTACCAAGCTCTGCAATTCATGAGTGGATAGAGAAACCG 726
   ::::::::::::::::::::
240 oserValIytrHisIySerLeuIleIySerIleMetIyIyIySaIuM 257
   ::::::::::::::::::::
727 TCCAG 731
   ::::::::::::::::::::
257 etIyS 258
   ::::::::::::::::::::
seq_name: SwissProt_37:TRY3_RAT

```

## seq\_documentation\_block:

ID TRY3\_RAT STANDARD: PRT: 247 AA.

AC P08426:

DT 01-AUG-1988 (REL. 08, CREATED)

DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)

DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)

DE TRYPsinogen III, CATIONIC PRECURSOR (EC 3.4.21.4) (PRETRYPsinogen III).

OS RATUS NORVEGICUS (RAT).

OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; RATTUS.

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 87271609.

RA FLETCHER T.S., ALHADEFF M., CRAIK C.S., LARGMAN C.;

RT Isolation and characterization of a cDNA encoding rat cationic trypsinogen.

RL BIOCHEMISTRY 26:3081-3086(1987).

CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.

CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1, ALSO KNOWN AS THE TRYPsin FAMILY.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----

CC EMBL: M16624; G206499; -.

DR PIR: A27547; A27547.

DR PROSITE: PS00134; TRYPsin\_HIS. 1.

DR PROSITE: PS00135; TRYPsin\_SER. 1.

DR PFM: PF00089; trypsin. 1.

DR HSP: P00763; IDPO.

KW HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZMOGEN; SIGNAL;

KW MULTIGENE FAMILY.

FT SIGNAL 1 15

FT PROPEP 16 24

FT CHAIN 25 247

FT ACT\_SITE 64 64

FT ACT\_SITE 108 108

FT ACT\_SITE 201 201

FT DISULFID 31 161

FT DISULFID 49 65

FT DISULFID 133 234

FT DISULFID 140 207

FT DISULFID 172 186

FT DISULFID 197 221

FT STPE 195 195

SQ SEQUENCE 247 AA: 26269 MW; 7DD35F0B CRC32;

## alignment\_scores:

Quality: 452.00 Length: 243

Ratio: 2.723 Gaps: 6

Percent Similarity: 68.313 Percent Identity: 37.860

## alignment\_block:

US-09-030-606-173 x TRY3\_RAT

Align seg 1/1 to: TRY3\_RAT from: 1 to: 247

```

10 CACTCGCAGCCCTGGCAGGCGGACATGTCATGGAAGAAAGAAATGTTCTG 59
   ::::::::::::::::::::::::::::
34 AsnserleuprIytrGlnValSerleu...AsnAlaGlyTrHisphCys 49
   ::::::::::::::::::::::::::::::
60 CTGCGGCGTCTGTCATCGCAGTGGGTCTGTGTCGAGCGGACACTGTT 109
   ::::::::::::::::::::::::::::::
49 sgllyglSerleuIleAsnserGlnItrpValValSerAlaIleHisCysT 66
   ::::::::::::::::::::::::::::::

```

```

110 TCAGAACCTCTACACCATCGGCTGGCTGACAGTCTTGAGCCGAC 159
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
66  yrlsserargileglnvalargleuglgnlhlslasnlalaspval... 81
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
160 CAGAGCCAGGAGGACCATGATGAGAGCCAGCTCTCCGACGCGACC 209
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
82  valgluglglglglgnlgnlphlelaspvalalalaysllelglarghlsp 98
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
210 AGACTACACAGACCCCTGCTGCTGCTACGACCTCATGCTCAAGTTGG 259
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
98  osettyrhslalnlnthrpheaspasnpillemetleuileylslslea 115
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
260 ACGAATCCGCTCCGAGTGTACACATCCGAGCATCAGCATGCTTCG 309
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
115 snserprolathrlleuasnserarvgvalserthrvalserleuproarg 131
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
310 CAGTGGCCCTACCGGCGGAGACTCTTGCTCTTCTGCTGGGCTTCCT 359
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
132 sercysglsersergltythrlyscysleuvalsergltyrpgly.... 146
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
360 GCGCAGCGGTAGCTCAGCGGTGTGTCTGCTCCCTCTCAAGAGGTGCT 409
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
147  ....  :::::  :::::  :::::  :::::  :::::  :::::  :::::
410 CTGCCAGTCCGCGGGGCTGACCCAGACCTGCTGCCAGCAGAAATGC 459
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
148 hrlserser..glythrash.....  :::::  :::::  :::::  :::::  :::::  :::::  :::::
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
460 CTACCGCTCTCAGTGTGCTGACAGTGTGCTGCTGCTGAGAGAGTCTGC 509
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
156 rosetleuileuglncylsleuaspalaprovalleuaseraspsercys 172
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
510 AGTAGGCTATGACCCGCTGTACACCCAGCAGCTGTCTGCGCGCGGG 559
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
173 lyserseertyrprogllylsiliethrserasnecpescysleuaglyph 189
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
560 AGGCAAGACAGCAGAGAGACTCTGCAACGCTGAGCTGGGGGCGCCCTGA 609
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
189 eleuglncylgllylsaspsercysglnlyaspsercylgllyprovaly 206
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
610 TCTGCAACGGGACTCTGCAAGGCGCTGTGCTTTTCGAAAGCCCGCTGT 659
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
206 alcyasnglyglleuileuglncylvalalsertrpglytyr-gly... 221
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
660 GGCAGAGTGTGGCGTGCAGTGTCTACACCAACCTGCAATATCAGCA 709
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
222 Alaelnlysglylsyprogllyvaltyrthrlyvalcysasnlyrvalas 238
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
710 GTGATAGAGAAACCGTCCAGCCAGT 737
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
238 ntrpilleglnlnthrlvalalalaspasn 247
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
seq_name: Swlssprot_37:KLKL_MOUSE
seq_documentation_block:
ID  KLKL_MOUSE  STANDARD:  PRT:  259 AA.
AC  P15948:
DT  01-APR-1990 (REL. 14, CREATED)
DT  01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT  01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE  GLANDULAR KALLIKREIN K22 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE  (MGK-22) (EPIDERMAL GROWTH FACTOR-BINDING PROTEIN TYPE A) (EGF-BP A)
DE  (NERVE GROWTH FACTOR BETA CHAIN ENDOPEPTIDASE) (BETA-NGF-
DE  ENDOPEPTIDASE).
GN  KLK22 OR KLK-22.
OS  MUS MUSCULUS (MOUSE).
OC  EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC  RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN-BALB/C; TISSUE-SALIVARY GLAND;
RX  MEDLINE: 88107594.
RA  DRINKWATER C.C., EVANS B.A., RICHARDS R.I.;

```

```

RT  "Mouse glandular kallikrein genes: identification and
RT  characterization of the genes encoding the epidermal growth factor
RT  binding proteins.";
RT  BIOCHEMISTRY 26:6750-6756(1987).
RN  [2]
RN  SEQUENCE OF 25-54.
RX  MEDLINE: 91190897.
RA  FAHNESTOCK M., WOO J.E., LOPEZ G.A., SNOW J., WALZ D.A., ARICI M.J.,
RA  MOBLEY W.C.;
RT  "Beta-NGF endopeptidase: structure and activity of a kallikrein
RT  encoded by the gene mgk-22.";
RT  BIOCHEMISTRY 30:3443-3450(1991).
RN  [3]
RN  SEQUENCE OF 25-41.
RC  TISSUE-SUBMANDIBULAR GLAND;
RX  MEDLINE: 92348355.
RA  PETERS J., TAKAHASHI S., TADA M., MIYAKE Y.;
RT  "MGK-6-derived true tissue kallikrein is synthesized, processed, and
RT  targeted through a regulated secretory pathway in mouse pituitary
RT  AtT-20 cells.";
RL  J. BIOCHEM. 111:643-648(1992).
RN  [4]
RN  SEQUENCE OF 17-54 AND 70-120 FROM N.A.
RX  MEDLINE: 87250386.
RA  EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;
RT  "Mouse glandular kallikrein genes. Structure and partial sequence
RT  analysis of the kallikrein gene locus.";
RL  J. BIOL. CHEM. 262:8027-8034(1987).
CC  -1- FUNCTION: GLANDULAR KALLIKREIN CLEAVE MET-LYS AND ARG-SER BONDS
CC  IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC  -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
CC  SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC  KALLIDIN (LYSLYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC  MET-1-XAA OR LEU-1-XAA.
CC  -1- SIMILARITY: BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC  TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: M17979; G387167; ALT-SEQ.
DR  EMBL: M17977; G387167; JOINED.
DR  EMBL: M17978; G387167; JOINED.
DR  EMBL: M18598; G198560; -.
DR  EMBL: M18618; G198561; -.
DR  PIR: A29746; A29746.
DR  PIR: A38356; A38356.
DR  PROSITE: PS00134; TRYPsin_HIS, 1.
DR  PROSITE: PS00135; TRYPsin_SER, 1.
DR  PFM: PF00089; trypsin, 1.
DR  HSSP: P00757; ISGF.
KM  MULTIGENE FAMILY; ZIMOGEN; SIGNAL.
KM  MULTIGENE FAMILY; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
FT  SIGNAL 1 17
FT  PROPEP 18 24
FT  CHAIN 25 259
FT  ACT_SITE 65 65
FT  ACT_SITE 118 118
FT  ACT_SITE 211 211
FT  DISULFID 31 171
FT  DISULFID 50 66
FT  DISULFID 150 217
FT  DISULFID 182 196
FT  DISULFID 207 232
FT  CARBOHYD 102 102
SQ  SEQUENCE 259 AA; 28384 MW; 3B02165A CRC32;

```

## alignment\_scores:

Quality: 452.00 Length: 251  
 Ratio: 2.790 Gaps: 4  
 Percent Similarity: 64.542 Percent Identity: 38.247

## alignment\_block:

US-09-030-606-173 x KLK3\_MOUSE ..

Align seg 1/1 to: KLK3\_MOUSE from: 1 to: 259

```

10 CACTCGACGCGCTGGGAGGCGGCGACCTGTATGGAACAAATTGTTCTG 59
   : : : : : : : : : : : : : : : : : : : : : : : : : :
34 AAsbserGlnProtrpGlnValAlaValTyrTylLeuAspGlnTyrLeu 50
   : : : : : : : : : : : : : : : : : : : : : : : : : :
50 sGlyGlyValLeuLeuAspArgAsnTrpValLeuThrAlaAlaHisCys 67
   : : : : : : : : : : : : : : : : : : : : : : : : : :
110 TCCAGAACTCTACACCATCGGCGGCTGGCGCTGCACAGCTTTCAGGCCGAC 159
   : : : : : : : : : : : : : : : : : : : : : : : : : :
67 YrGlnAspLysTyrAsnIleTrpLeuLysAsnLysLeuPheGlnAsp 83
   : : : : : : : : : : : : : : : : : : : : : : : : : :
160 CAAGAGCCAGGAGCCAGATGGTGGAGCCAGCCTTCCTCAGCGCACCC 209
   : : : : : : : : : : : : : : : : : : : : : : : : : :
84 ...GlnProSerAlaGlnHisArgLeuValSerLysSerPheProHisPr 99
   : : : : : : : : : : : : : : : : : : : : : : : : : :
210 AGAGTACACAGACCCCTTG.....CTCG 232
   : : : : : : : : : : : : : : : : : : : : : : : : : :
99 CAAPPhasAsnMetSerLeuLeuGlnSerValProThrGlyAlaAspLeu 116
   : : : : : : : : : : : : : : : : : : : : : : : : : :
233 CTACAGCACTCATGCTCATCAAGTTGGAGCAATCCGTGCTCCAGTCTGAC 282
   : : : : : : : : : : : : : : : : : : : : : : : : : :
116 eArAsnAspLeuMetLeuLeuArgLeuSerLysProAlaAspIleThrAsp 132
   : : : : : : : : : : : : : : : : : : : : : : : : : :
283 ACATCCGGAGCATCAGCATGCTGCCACTGCCCTACCCGGGGAATC 332
   : : : : : : : : : : : : : : : : : : : : : : : : : :
133 ValValLysProIleAspLeuProThrGlnProLysLeuGlySerTh 149
   : : : : : : : : : : : : : : : : : : : : : : : : : :
333 TTGCTCTGTTGGTGGGCTGCTGCTGGCGAAGCGGTGAGCTCAGCGGTG 382
   : : : : : : : : : : : : : : : : : : : : : : : : : :
149 rCySLeuAlaSerGlyTrpLys..... 156
   : : : : : : : : : : : : : : : : : : : : : : : : : :
383 TGTGTGTCCTCTTCAAGAGGTCTCTGCGCAGTGGCGGGGCTGACC 432
   : : : : : : : : : : : : : : : : : : : : : : : : : :
156 ..... 156
   : : : : : : : : : : : : : : : : : : : : : : : : : :
433 CAGAGCTCTGCTCCAGGC.....AGAATGCTTACCTGCTGCGCATG 475
   : : : : : : : : : : : : : : : : : : : : : : : : : :
157 .....SerIleAsnGlnLeuIleTyrGlnAsnProAsnAspLeuGlnCy 171
   : : : : : : : : : : : : : : : : : : : : : : : : : :
476 CGTGAACGTGTCGGTGTCTGAGAGAGTCTGCAGTAAGCTCATGACC 525
   : : : : : : : : : : : : : : : : : : : : : : : : : :
171 sValSerIleLysLeuHisProAsnGlnValCysValLysAlaHisIleL 188
   : : : : : : : : : : : : : : : : : : : : : : : : : :
526 CGCTGTACCAACCCAGCATGTTCTGCCCGCGGAGGAGGCAAGACCAAG 575
   : : : : : : : : : : : : : : : : : : : : : : : : : :
188 eulysValThrAspValMetLeuCysAlaGlyIleuMetAsnGlyLys 204
   : : : : : : : : : : : : : : : : : : : : : : : : : :
576 GATCTCCGCAACGCTGACTGTGGGGGCCCTGATCTGCACAGGGTCT 625
   : : : : : : : : : : : : : : : : : : : : : : : : : :
205 AspThrCysLysSerGlyAspSerGlyLysProLeuLysAspGlyValLe 221
   : : : : : : : : : : : : : : : : : : : : : : : : : :
626 GCAAGGCGCTGTCTTTCGAAAAAGCCGCTGGCCAGTGGCGGCTGC 675
   : : : : : : : : : : : : : : : : : : : : : : : : : :
221 uGlnGlyIleThrSerTrpLysThrProCysGlyGlnProAsnAlaP 238
   : : : : : : : : : : : : : : : : : : : : : : : : : :
676 CAGGTGTCTACCAACCTCTGCAATTCAGTGAAGTGAAGAAAAAC 725
   : : : : : : : : : : : : : : : : : : : : : : : : : :
238 rAlaIleTyrThrLysLeuIleLysSerThrSerTrpLysAspThr 254
   : : : : : : : : : : : : : : : : : : : : : : : : : :
726 GTC 728
   : : :
255 Web 255

```

seq\_name: SwissProt\_37:KLK3\_MOUSE

seq\_documentation\_block:

```

ID KLK3_MOUSE STANDARD: PRT: 261 AA.
AC P00756;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN K3 PRECURSOR (EC 3.4.21.35) (TISSUE KALLIKREIN)
DE (MGK-3) (7S NERVE GROWTH FACTOR GAMMA CHAIN) (GAMMA-NGF).
GN KLK3 OR KLK-3 OR NGGF.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MORIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85076169.
RA ULLRICH A., GRAY A., WOOD W.I., HAYFLICK J., SEEBURG P.H.;
RT "Isolation of a cDNA clone coding for the gamma-subunit of mouse
RT nerve growth factor using a high-stringency selection procedure.";
RL DNA 3:387-392(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85257431.
RA EVANS B.A., RICHARDS R.I.;
RT "Genes for the alpha and gamma subunits of mouse nerve growth factor
RT are contiguous.";
RL Embo J. 4:113-118(1985).
RN [3]
RP SEQUENCE OF 25-261.
RX MEDLINE: 81264363.
RA THOMAS K.A., BAGIAN N.C., BRADSHAW R.A.;
RT "The amino acid sequence of the gamma-subunit of mouse submaxillary
RT gland 7 S nerve growth factor.";
RL J. Biol. Chem. 256:9156-9166(1981).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3.15 ANGSTROMS) OF 7S COMPLEX.
RX STRAIN-SWISS WEBSTER; TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE: 98035451.
RA BAX B., BLONDEL T.L., MURRAY-RUST J., McDONALD N.O.;
RT "Structure of mouse 7S NGF: a complex of nerve growth factor with
RT four binding proteins.";
RL STRUCTURE 5:1275-1285(1997).
CC -!- FUNCTION: 7S NGF ALPHA CHAIN STABILIZES THE 7S COMPLEX. THE BETA
CC DIMER PROMOTES NEURITE GROWTH. THE GAMMA CHAIN IS AN ARGININE-
CC SPECIFIC PROTEASE; IT MAY ALSO HAVE PLASMINOGEN ACTIVATOR
CC ACTIVITY, AS WELL AS MITOGENIC ACTIVITY FOR CHICK EMBRYO
CC FIBROBLASTS.
CC -!- SUBUNIT: 7S NERVE GROWTH FACTOR IS COMPOSED OF TWO ALPHA CHAINS,
CC A BETA DIMER COMPOSED OF IDENTICAL CHAINS, AND TWO GAMMA CHAINS.
CC -!- THIS PRECURSOR IS CLEAVED INTO SEGMENTS TO PRODUCE THE ACTIVE
CC FORM OF THE GAMMA CHAIN, WHICH OCCURS NATURALLY AS COMBINATIONS
CC OF EITHER TWO OR THREE SEGMENTS HELD TOGETHER BY DISULFIDE
CC BONDS: B1 + A OR B1 + C + B2.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as there are no content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: X01389; G53374; -.
CC EMBL: X01798; G582429; -.
CC EMBL: X01799; E73575; -.
CC PIR: A00942; NGMSG.
CC PDB: JSGF; 27-MAY-98.
CC MGD: MGI:97322; NGFG.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.

```







FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 152 219 BY SIMILARITY.  
 FT DISULFID 184 198 BY SIMILARITY.  
 FT DISULFID 209 234 BY SIMILARITY.  
 FT CARBOHYD 102 102 POTENTIAL.  
 SQ SEQUENCE 261 AA; 28671 MW; 99CF6275 CRC32;

## alignment\_scores:

Quality: 449.50 Length: 254  
 Ratio: 2.863 Gaps: 5  
 Percent similarity: 61.811 Percent identity: 37.402

## alignment\_block:

US-09-030-606-173 x KLR2\_HUMAN ..

Align seg 1/1 to: KLR2\_HUMAN from: 1 to: 261

```

10 CACTGCAGCCCTGGCGAGCGGACATGTCATGAAACGAATTTGTTCTG 59
|||||
34 HISSERGLNPROTIRPGLNVALALAVALLYRSEHISGLYTRPALAHISCY 50
CTGGGCGCTGCTGTGCATCCGACAGTGGTGTCTGACGGCAGACTGT 109
|||||
50 sclyglvalleuvalhisproglntprvalleuthralalahlscysl 67
110 TCCAGAACCTCCTACACCATCGGCGCTGGCTGCACAGTTTGAGCCGAC 159
|||||
67 eulylalysalsergluvaltrleuglyarhlslasnlueupheglupro 83
160 CAAGAGCCGAGGAGCGATGTGGAGCGACCTCTCCGTACGGCAGCC 209
|||||
84 gluaaprhicly...gluarlyalprovalserhlaserphrohispr 99
210 AGAGTACACAGACCCCTGCTC..... 231
99 oleuylrasmetserleuylshisglinslerleuargproahpqlua 116
232 ....GCTAACGACCTCATGCTCATCAATTTGAGCAATCCCTGTCGAG 276
|||||
116 apseraserlslaspleumelleuleuargleuaserluproahlysl 132
277 TCTGACACCATCCGAGCATGACATGCTGTCGACGCTTACCGCGGG 326
|||||
133 thrsapvalvallylvalleuylleuprothgluproahleucl 149
327 GAACCTTTCGCTGTTCTGCTGGGCTGCTGCGGAAGGCTGACGTCA 376
|||||
149 ythrthrcysylalaserglytrpilyserile...glu..... 161
377 CGGCTGTGTCTGCTGCTTCAAGAGGCTCTCTCCAGTCCGCGGG 426
161 ..... 161
427 CTGACCCAGAGCTGTGCTGCCAGGAGATGCGTACGCTGTCAGTGC 476
|||||
162 .....ArgProArgSerleuaglnCys 173
477 GTGAAGCTGCTGCTGCTGCTGAGAGGCTGCTGACAGTCTATGACCC 526
|||||
174 valserleuhlslleuenserlnaspmetyalalargalatysergl 190
527 GCTGTACACCCAGCATGTTCTGCGCGCGGAGGAGGACAGACAGAG 576
|||||
190 ulysvaltrhgluphemetleucysalagllyleuprthrglygllysa 207
577 ACTGCTGCAAGGTGACTGTGGGCGCCCTGATGTCAGCGGCTGCTG 626
|||||
207 spthrcysgllylaspserglygllyproleuvalcysasnlyalbeu 223
627 CAGGGCTGTGTCTTTTCGAAAGCCCGGTGTGCTGCAAGTTGGCTGTC 676
|||||
224 glnglyllethrsertrpilyprogluprocysalaleuproglulyspr 240

```

677 AGGTGTCACACCACTGTCGAATTCATGAGTGGATAGAAACCG 726  
 |||||  
 240 calavalllytrhlylvalvalhlstlyrarglystrpilleysasprhri 257  
 727 TCCAGCCAGT 737  
 ::|||  
 257 lealalalasn 260

seq\_name: SwissProt\_37:KLR7\_RAT

seq\_documentation\_block: PRT; 261 AA.

```

ID KLR7_RAT STANDARD;
AC P36373;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN 7, SUBMANDIBULAR/RENAL PRECURSOR (EC 3.4.21.35)
DE (TISSUE KALLIKREIN) (RKG-7) (ESTERASE B) (PROTEINASE A).
GN KLR7 OR KLR-7
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RA CHEN Y.-P., CHAO J., CHAO L.;
RT "Molecular cloning and characterization of two rat renal kallikrein
genes."
RL BIOCHEMISTRY 27:7189-7196(1988).
RN [2]
RP SEQUENCE OF 25-75.
RC TISSUE-SUBMAXILLARY GLAND;
RX MEDLINE; 88198057.
RA KATO H., NAKANISHI E., ENYOJI K., HAYASHI I., OH-ISHI S., IWANAGA S.;
RT "Characterization of serine proteinases isolated from rat
submaxillary gland: with special reference to the degradation of rat
kinogens by these enzymes."
RL J. BIOCHEM. 102:1389-1404(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-|-XAA BONDS IN
SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
KALLIDIN (LYSL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
MET-|-XAA OR LEU-|-XAA.
CC -1- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND.
CC -1- SIMILARITY: BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC TRYPsin FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
entitles requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M19647; G205000; -
DR PIR; A31136; A31136.
DR PIR; B41429; B41429.
DR PROSITE; PS00134; TRYPsin_HIS. 1.
DR PROSITE; PS00135; TRYPsin_SER. 1.
DR PRAM; P00089; trypsin. 1.
DR HSSP; P00759; ITON.
KW HYDROLASE; SERINE PROTEASE; KININOGENASE; GLYCOPROTEIN;
KW MULTIGENE FAMILY; ZMOGEN; SIGNAL.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 261 GLANDULAR KALLIKREIN 7.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173 BY SIMILARITY.

```

FT DISULFID 50 66 BY SIMILARITY.  
 FT DISULFID 152 219 BY SIMILARITY.  
 FT DISULFID 184 198 BY SIMILARITY.  
 FT DISULFID 209 234 BY SIMILARITY.  
 FT CARBOHYD 108 108 POTENTIAL.  
 FT CONFLICT 35 35 S -> D (IN REF. 2).  
 FT CONFLICT 46 46 T -> S (IN REF. 2).  
 SQ SEQUENCE 261 AA: 28972 MW: 2785842 CRC32:

alignment\_scores:  
 quality: 449.00 length: 255  
 ratio: 2.738 gaps: 5  
 Percent Similarity: 64.314 Percent Identity: 36.863

alignment\_block:  
 US-09-030-606-173 x KLU7\_RAT ..

Align seg 1/1 to: KLU7\_RAT from: 1 to: 261

```

10 CACTCGCAGCCCTGCGAGCGGCGGCTGCTATGGAAAAGCATTTGCTG 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
34 AsnSerGlnProTrpGlnValAlaLeuYrSerPheThrLysTyrLeuCy 50
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 CTCGCGGCTGCTGGTGCATCCGAGATGGGCTGTGACAGCCGACATGTT 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
50 SGlyGlyValLeuLeuAspProSerTrpValIleThrAlaIleAspCys 67
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
110 TCAGAACTCTACACATCGGCTGGGCTGCGACATGTTGAGCGGAC 159
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
67 eSerAsnAsnTyrGlnValTyrLeuGlyArgAsnAsnLeuGlnAsp 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
160 CAAGAGCCAGGAGCCAGATGTGGAGGCCAGCTCTCCGAGCGGACCC 209
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
84 ...GlnProPheAlaGlnIleArgLeuValSerGlnSerPheProHisPr 99
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
210 AGAGTAAACAGACCCCTGCTC..... 231
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
99 oAsPTr...LysProPheLeuMetArgAsnIleThrArgLysProGlyA 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
232 .....GCTAACGACCTCATGCTCATCAATTGAGGAAATCCGCTGC 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
115 sPAsrHisSerAsnAspLeuMetLeuHisLeuSerGlnProAlaAsp 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
274 GAGTGTGACACCATCCGAGATGACATGCTCTGCGACGCTCCATACCC 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
132 IleThrAspGlyValIleValIleAspLeuProThrGlnGluProLysVa 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
324 GGGGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
148 lGlySerThrCysLeuAlaSerGlyTyrGly..... 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 TCACGGGTGTGTCTGCTGCTCTTCAGAGAGGTCTGCTGCCAGTCCGG 423
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
158 ..... 158
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
424 GGGGTGACCCAGAGCTGCGCTCCAGGC.....AGATGCTTACCGT 466
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
159 .....SerThrLysProLeuIleTrpGlnPheProAspAs 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
467 GCTCAGTGCCTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 516
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 pLeuGlnCysValAsnIleHisLeuLeuSerAsnGluLysCysIleLysA 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
517 TCTATGACCCGCTGTATACCCAGCATGTTCTGCGCGGAGGAGGCA 566
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 lAtyLysGlnLysValIleThrAspLeuMetLeuCysAlaGlyIleGln 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
567 GACAGAAAGACTCCTGACAGGTGACCTCTGGGGGCCCCGATGATGAA 616
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
204 GlyLysLysAspTrnCysThrGlyAspSerGlyLysProLeuLeuCysAs 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
617 CGGTACTTGCAGGCGCTTGTCTTTCGAAAAGCCCCGCTGCGCCAA 666

```

```

220 pGlyValLeuGlnGlyIleThrSerTrpGlySerValProCysAlaLysT 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
667 TTGGCGTCCAGAGTGTCTACACCAACCTGTGAAATCTACTGAGTGAGA 716
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 hAsnMetProAlaIleTyrThrLysLeuIleLysPheThrSerTrpIle 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
717 GAGAAACCGCTCCAG 731
   : : : : : : : : : : : : : : : : : : : : : : : : : : :
254 LysGlnValMetLys 258
   : : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: SwIsProt_37:KLU7_MOUSE
seq_documentation_block:
ID KLU7_MOUSE STANDARD: PRT: 261 AA.
AC P15946:
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE GLANDULAR KALLIKREIN K11 PRECURSOR (EC 3.4.21.35), SUBMANDIBULAR
DE (TISSUE KALLIKREIN) (MGR-11).
GN KLU11 OR KLU-11.
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIROGNATHI; MURIDAE; MURINAE; MUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BAB/C; TISSUE-LIVER.
RX MEDLINE: 89083511.
RA DRINKWATER C.C., RICHARDS R.I.;
RT "Sequence of mGK-11, a mouse glandular kallikrein gene.";
RL NUCLEIC ACIDS RES. 16:10918-10918(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85257431.
RA EVANS B.A., RICHARDS R.I.;
RT "Genes for the alpha and gamma subunits of mouse nerve growth factor
are contiguous.";
RL EMBO J. 4:133-138(1985).
RN [3]
RP SEQUENCE OF 16-54 AND 69-122 FROM N.A.
RX MEDLINE: 87250386.
RA EVANS B.A., DRINKWATER C.C., RICHARDS R.I.;
RT "Mouse glandular kallikrein genes. Structure and partial sequence
analysis of the kallikrein gene locus.";
RL J. BIOL. CHEM. 262:8027-8034(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-1-XAA BONDS IN
SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
MET-1-XAA OR LEU-1-XAA.
CC -1- SIMILARITY: BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL: X13215; G53066; -
CC EMBL: X13216; G582395; ALT_SEQ.
CC EMBL: X13217; E14069; -
CC EMBL: X13217; E14070; -
CC EMBL: X13218; E7100; -
CC EMBL: M18590; G554179; ALT_SEQ.
CC EMBL: M18610; G198537; -
CC PIR: S01971; S01971.
CC MGD: MGI:892023; KLU11.
CC PROSITE: PS00134; TRYPSIN_HIS. 1.

```



RP SEQUENCE FROM N.A.  
RC TISSUE-PROSTATE;  
RX MEDLINE: 89165891.  
RA RIEGMAN P.H.J., VLIESTRA R.J., VAN DER KORPUT J.A.G.M., ROMJIN J.C.,  
RA TRAPMAN J.;  
RT "Characterization of the prostate-specific antigen gene: a novel  
RT human kallikrein-like gene.";  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 159:95-102(1989).  
RN [17]  
RP SEQUENCE FROM N.A.  
RC TISSUE-PROSTATE;  
RX MONNE M.M., MORENO J.M., MELE C.M., MULHOLLAND G.M., GOMELLA L.G.;  
RL SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [8]  
RP SEQUENCE OF 17-261 FROM N.A.  
RX MEDLINE: 88289366.  
RA SCHULZ P., STUCKA R., FEIDMANN H., COMBRATO G., KLOBECK H.-G.,  
RA FITTLER F.;  
RT "Sequence of a cDNA clone encompassing the complete mature human  
RT prostate specific antigen (PSA) and an unspliced leader sequence.";  
RL NUCLEIC ACIDS RES. 16:6226-6226(1988).  
RN [9]  
RP SEQUENCE OF 25-261.  
RX MEDLINE: 86205857.  
RA WATT K.W.K., LEE P.J., M'TINKULU T., CHAN W.P., LOOR R.;  
RT "Human prostate-specific antigen: structural and functional  
RT similarity with serine proteases.";  
RL PROC. NATL. ACADE. SCI. U.S.A. 83:3166-3170(1986).  
RN [10]  
RP SEQUENCE OF 25-261.  
RX MEDLINE: 88082806.  
RA SCHALLER J., AKIYAMA K., TSUDA R., HARA M., MARTI T., RICKLI E.E.;  
RT "Isolation, characterization and amino-acid sequence of gamma-  
RT semioprotein, a glycoprotein from human seminal plasma.";  
RL EUR. J. BIOCHEM. 170:111-120(1987).  
RN [11]  
RP 3D-STRUCTURE MODELLING.  
RX MEDLINE: 95218633.  
RA VILLOUTREIX B.O., GETZOFF E.D., GRIFFIN J.H.;  
RT "A structural model for the prostate disease marker, human prostate-  
RT specific antigen.";  
RL PROTEIN SCI. 3:2033-2044(1994).  
CC -1- FUNCTION: PRESUMABLY HYDROLYZE THE HIGH MOLECULAR MASS SEMINAL  
CC VESICLE PROTEIN THUS LEADING TO THE LIQUEFACTION OF THE SEMINAL  
CC COAGULUM.  
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: TYR-I-XAA.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: X14810; G296672; -  
DR EMBL: X13940; G35723; -  
DR EMBL: X13941; E14018; ALT\_SEQ.  
DR EMBL: X13942; E14019; -  
DR EMBL: X13943; E14020; -  
DR EMBL: X13944; E14022; -  
DR EMBL: X05332; G35741; -  
DR EMBL: X07730; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: M27274; G190553; -  
DR EMBL: M26663; G618464; -  
DR EMBL: U17040; G595946; -  
DR EMBL: M24543; G511857; ALT\_SEQ.  
DR PIR: S03604; S03604.  
DR PIR: A26757; A26757.  
DR PIR: A32297; A32297.  
DR PIR: A32423; A32423.

DR PDB: 1PFA; 26-JAN-95.  
DR MIM: 176820; -;  
DR PROSITE: PS00134; TRYPSIN\_HIS; 1.  
DR PROSITE: PS00135; TRYPSIN\_SER; 1.  
DR PFA: PF00089; trypsin; 1.  
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ANTIGEN; ZMOGEN; SIGNAL;  
KW 3D-STRUCTURE.  
FT SIGNAL 1 18  
FT PROPEP 19 24  
FT CHAIN 25 261  
FT ACT\_SITE 65 65  
FT ACT\_SITE 120 120  
FT ACT\_SITE 213 213  
FT CARBOHYD 69 69  
FT DISULFID 31 173  
FT DISULFID 50 66  
FT DISULFID 152 219  
FT DISULFID 184 198  
FT DISULFID 209 234  
FT CONFLICT 64 73  
FT CONFLICT 73 73  
FT CONFLICT 86 86  
FT CONFLICT 94 94  
FT CONFLICT 136 136  
FT CONFLICT 165 168  
FT CONFLICT 175 175  
FT CONFLICT 184 184  
FT CONFLICT 260 260  
SQ SEQUENCE 261 AA; 28741 MW; 4723B52E CRC32;  
  
alignment\_scores: 448.50 Length: 254  
Ratio: 2.803 Gaps: 5  
Percent Similarity: 62.992 Percent Identity: 37.402  
  
alignment\_block:  
US-09-030-606-173 x PROS\_HUMAN ..  
  
Align seg 1/1 to: PROS\_HUMAN from: 1 to: 261  
  
10 CACTCGACGCCCTGGCGAGCGCGCATGTCATGAGAAACGATTTGTCG 59  
|||||  
34 Hiserlnprtrprlnvallleuvalalaserlglyarglavalcy 50  
CTCGGCGCTCGTGGTCATCCGACGTGGTGTGTCAGCCGACGTCGT 109  
|||||  
50 sglyglyvallleuvalhisprglntrpyallleuthralalancys1 67  
110 TCGAGACCTCTACACGCGGCTGGCGCTGCACAGTCGTCAGCGGAC 159  
|||||  
67 leagvalnlysservalilleuvaldeucllyarghisserleuphenispro 83  
160 CAAGAGCCGAGGAGCCAGATGTCGAGGCGCGCTCCCTACGCGGCC 209  
|||||  
84 gluasprhrcly...glvalrhneglvalserhisserphrohispr 99  
210 AGAGTACACAGACCCCTGCTCGCTAAC..... 237  
|||||  
99 oleutyraspmetserserleuvalysasnargphleuargproglyaspra 116  
116 sperserhisasprleuvalleuvaldeuarglsuargproalaglleu 132  
277 TCTGACACATCCGAGCATGTCGTCGAGCGCTCCAGTCCGCGGG 326  
|||||  
133 Thrspalavallysvalmetasprleuprothrglncglproalaleucl 149  
327 GACCTCTGCTCGTTTGTGCTGGGCTGCTGCGGACGCTGACCTCA 376  
|||||  
149 ythrthrcystylalaserglytrpglyserile...glu..... 161

```

377 CGGCTGTGTGTGCTCCCTCTTCAAGAGGTCTCTGCGCCAGTGGGGGG 426
161 ..... 161
427 CTGACCCAGACCTGTGGCTCCAGCAGATGCTTACCGTCTGTCAGTGC 476
162 .....ProgluLupheleu.....ThProLysLysLeuGlnCys 173
477 GTGACGTGTGGGTGTGTGAGAGGTCTGCAAGTAACTCTTATGACCC 526
174 ValAspLeuHisValIleSerAsnAspValLysAlaGlnValHisProG1 190
527 GCTGTACACCCAGCATGTCTGTGGCGCGGAGGGGCAAGACCAAGAG 576
190 nlvsvalThrlsPhemeLleuCysAlaGlyArgTrpThrLysLys 207
577 ACTCTGCAACGCTGACTCTGTGGGGGCGCCCTGATGCAACGGTACTTG 626
207 erThrcysserGlyAspserGlyProLeuValCysAsnGlyValLeu 223
627 CAGGCGCTGTGTCTTTCGAAAGCCCGGTGTGCGCAAGTTGGCGTCC 676
224 GlnGlyIleThrSerTrpGlySerGluProCysAlaLeuProGluArgPr 240
677 AGGTGTACACGCTGCTGCAATTCACGTGAGTGCATAGAGAAACCG 726
240 oSerLeuThrLysValValHisTyrArgLysTrpLysAspTrp 257
727 TCAGAGCCAGT 737
257 lvalAlaAsn 260

seq_name: SwissProt_377.TRYB_RAT

seq_documentation_block:
ID TRYB_RAT STANDARD: PRT: 246 AA.
AC P32822:
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE TRYPsinogen v-b precursor (EC 3.4.21.4).
OS RATUS NORVEGICUS (RAT).
NC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-PANCREAS;
RX MEDLINE: 92165057.
RA KANG J., WIEGAND U., MUELLER-HILL B.;
RT Identification of cDNAs encoding two novel rat pancreatic serine
RT proteases.
RL GENE 110:181-187(1992).
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC .....
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC .....
CC EMBL: X59013: G57415: -
CC PIR: J01472: J01472.
CC PROSITE: PS00135: TRYPSIN_SER: FALSE-NEG.
CC PROSITE: PS00134: TRYPSIN_HIS: 1.
CC PFAM: PF00089: trypsin: 1.
CC HSSP: P00763: IDPO.
CC HYDROLASE; SERINE PROTEASE; DIGESTION; PANCREAS; ZYMOGEN; SIGNAL;
KM

```

```

KM MULTIGENE FAMILY.
FT SIGNAL 1 15
FT PROPEP 16 24
FT CHAIN 25 246
FT ACT_SITE 64 64
FT ACT_SITE 108 108
FT ACT_SITE 200 200
FT DISULFID 31 160
FT DISULFID 49 65
FT DISULFID 133 233
FT DISULFID 140 206
FT DISULFID 171 185
FT DISULFID 196 220
FT SITE 194 194
SO SEQUENCE 246 AA; 26819 MW; 743C84B CRC32;

alignment_scores:
Quality: 448.00 Length: 243
Ratio: 2.835 Gaps: 4
Percent Similarity: 65.021 Percent Identity: 38.683

alignment_block:
US-09-030-606-173 x TRYB_RAT ..

Align seg 1/1 to: TRYB_RAT from: 1 to: 246

10 CACTCGAGCCCTGGAGCGGCGCAGTGTGATGGAAGAAAGATTTCTG 59
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 HisSerValProTyrGlnValSerLeuAsnAlaGlySerHisIle...Cy 49
60 CTGCGGGCTGCTGTGTCATCCGAGTGGTGTCTGACGGCGACACTGT 109
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
49 sGlyGlySerLeuIleThrAspGlnTrpValLeuSerAlaAlaHisCys 66
110 TCAGACTCTCTACACCATCGGCTGGGCTGCACAGTCTGAGCGGAC 159
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
66 yHisProGlnLeuGlnValArgLeuGlyGlnHisAsnIle...TyrGlu 81
160 CAGAGCGGAGGAGCCAGATGTGTGAGCGCAGCTCTCCGTACGGCACC 209
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 IleGlnGlyAlaGlnGlnPheIleAspAlaAlaLysMetIleLeuHisPr 98
210 AGATACACAGACCTTCTCCCTAACGACCTCATCTCATCAAGTTGG 259
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 oAspTyrAspLysTrpThrValAspAsnAspIleMetLeuLeuLysLeu 115
260 AGCAATCGGTGTGAGTCTGACACACATCCGAGACATCAGCATGCTTGC 309
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 ySerProAlaThrLeuAsnSerLysValSerThrIleProLeuProGln 131
310 CAGTCCCTACCGCGGGGAGACTTGTGCTGTTCTGCTGGGGTCTGCT 359
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 TyrLysProThrAlaGlyThrGlnCysLeuValSerGlyTrpGlyVal 148
360 GCGCAAGGTGAGCTCACGGGTGTGTGCTGCTTCAAGAGGTCT 409
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 nlvsPheGly..... 151
410 CTGCCAGTGGCGGGGCTGACCCAGAGCTGTGCTCCAGCAGATG. 458
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
152 .....PheGlnSer 154
459 CTAACGTGTCTGAGTGCAGTGCAGTGTGCTGCTGAGAGGTCTG 508
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
155 ProSerValLeuGlnCysLeuAspAlaProValLeuSerAspSerValCy 171
509 CAGTAAGCTCATGACCCGCTGTACACCCAGCATTTCTGCGCCGCG 558
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 HisLysAlaLysAlaProArgGlnIleThrAsnAsnMetPheCysLeuLysP 188
559 GAGGCAAGACCAAGAGACTCTGCAAGGTGACTGTGGGGGCGCCCTG 608
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

188 heLeuGlugLyLyAspSerCysGlnTyrAspSerGlyGlyProVal 204
609 ATCTGCAAGGACTTCGAGGGCTGTCTTTCGAAAGCCCGCTG 658
205 ValCysAsnGlyValGlnGlyIleValSerTyrGlySpgly...Cy 220
659 TGCCCAAGTTGGCGTCCAGGTGTACACCAACCTTCGAAATTCAGTG 708
220 sAlaLeuGlugLyLysProGlyValTyrThrLysValCysAsnTyrLeuA 237
709 ACTGATAGAGAAACCGTCCAGGCAGT 737
237 snTyrIleGlnGlnThrValAlaAlaAsn 246

seq_name: SwIsprProf_37:PSS9_HUMAN
seq_documentation_block:
ID PSS9_HUMAN STANDARD; PRT; 244 AA.
AC 092876;
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE PROTEASE M PRECURSOR (EC 3.4.21.-) (NEURONIN) (ZYME) (SP59).
GN PRSS9.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE: 97053999.
RA ANISOWICZ A., SOTIROPOULOU G., STENMAN G., MOK S.C., SAGER R.;
RT "A novel protease homolog differentially expressed in breast and
RT ovarian cancer.";
RL MOL. MED. 2:624-636(1996).
RL [2]
RP SEQUENCE FROM N.A.
RA TISSUE-COLON;
RC MEDLINE: 97157069.
RA YAMASHIRO K., TSURUOKA N., KODAMA S., TSUJIMOTO M., YAMAMURA Y.;
RA TANAKA T., NAKAZATO H., YAMAGUCHI N.;
RT "Molecular cloning of a novel trypsin-like serine protease (neurosin)
RT preferentially expressed in brain.";
RL BIOCHIM. BIOPHYS. ACTA 1350:11-14(1997).
RL [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 97460104.
RX LITTLE S.P., DIXON E.P., NORRIS F., BUCKLEY W., BECKER G.W.,
RA JOHNSON M., DOBBINS J.R., WYRICK T., MILLER J.R., MACKELLAR W.,
RA HEPBURN D., CORVALAN J., MCCLEURE D., LIU X., STEPHENSON D.,
RA CLEMENS J., JOHNSTONE E.M.;
RT "Zyyme, a novel and potentially amyloidogenic enzyme cDNA isolated
RT from Alzheimer's disease brain.";
RL J. BIOL. CHEM. 272:25135-25142(1997).
RL [4]
RP SUBCELLULAR LOCATION: SECRETED.
CC [1] TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN BRAIN. ALSO FOUND
CC IN COLON AND KIDNEY.
CC [2] SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
CC [3] This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [4] EMBL: U62801: G1518788; -
CC EMBL: D78203: G1805493; -
CC EMBL: AF013988: G2318115; -
CC MIM: 602652; -
CC [5] PROSITE: PS00134; TRYPSIN_HIS; 1.

```

```

DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR PFM: PF00089; trypsin; 1.
KW HYDROLASE; SERINE PROTEASE; GLYCOPROTEIN; ZYMOGEN; SIGNAL.
FT SIGNAL 1 16
FT PROPEP 17 21
FT CHAIN 22 244
FT ACT_SITE 62 62
FT ACT_SITE 106 106
FT ACT_SITE 197 197
FT DISULFID 28 157
FT DISULFID 47 63
FT DISULFID 131 231
FT DISULFID 138 203
FT DISULFID 168 182
FT DISULFID 193 218
FT CARBOHYD 134 134
SQ SEQUENCE 244 AA; 26856 MW; 0CFCEFC CRC32;

alignment_scores:
Quality: 448.00 Length: 244
Ratio: 2.835 Gaps: 5
Percent Similarity: 64.754 Percent Identity: 38.934

alignment_block:
US-09-030-606-173 x PSS9_HUMAN ..
Align seg 1/1 to: PSS9_HUMAN from: 1 to: 244

13 TGCGAGCCCTGGAGCGGCGGACGTGTCATGAAAGAAATGTTGTCG 62
32 SerHisProGlycIlnAlaIleuTyrThrSerGlyHisLeuLeuGly 48
63 GGGCGTCTGGTGCATCCGCGAGTGGGTGTCAGCGGACGATGTTCC 112
48 yGlyValIleuIleHisProLeuTyrValLeuThrAlaIlnHisCys 65
113 AGAACTCCTAACACACGCGGCTGGCGCTGCACAGTCTT.....GAG 153
65 ySerAsnLeuGlnValIlePheLeuGlyLysHisAsnLeuArgIlnArg 81
154 GCGGACCAAGAGCGGAGCGGACGAGATGTGTGAGCGGCGCTCCGTAG 203
82 SerSerGlnGlnGln...SerSerValAlaArgAla.....ValI 94
204 GCACCCAGAGTACACAGACACCTTCTCCCTAACACACCTATGTCATCA 253
94 eHisProAspTyrAspAlaIleSerHisAspGlnAspIleMetLeuA 111
254 AGTTGAGCGAATCCGTCGCGAGTGCACACCATCCGGAGCATACAGCAT 303
111 rGlyLeuAlaArgProAlaLysLeuSerGlyLeuIleGlnProLeuPro 127
304 GCTTCGAGTCCGCTACCGCGGGGAACCTTCGCTGTTCTGCTGGGG 353
128 GluArgAspCysSerAlaAsnThrThrSerCysHisIleLeuGlyTyr 144
354 TCTGCTGGGGAACGGTACGCTACGGGCTGTGTCTGCCCTTCAAGA 403
144 y.....
404 GGTCTCTGGCCAGTGGCGGGGCTGACCCAGACGCTGCTGCCAGCA 453
145 .....LysThrAlaAsp...GlyA 150
454 GAATGCTACCGTCTGACGAGTGCATGCTGCTGCTGCTGCTGAGAG 503
150 sPheProAspThrIleGlnCysAlaTyrIleHisLeuValSerArgGly 166
504 GTTGCAGTAAGCTATGACCCGCGGTGACACCCAGCATGTTCTGGCC 553
167 GluCysGlnHisAlaTyrGProGlyGlnIleThrGlnAsnMetLeuCysAl 183

```

```
554 CGCCGAGGGCAAGACCAAGAGACTCCTGCACCGTGACTCTGGGGGC 603
    |||  ::  |||
183 aGlyaspGluLysTyrGlyLysAspSerCysGlnGlyAspSerGlyGlyP 200
    |||
604 CCTGATCTGCACCGGTACTTGCAGGGCCTTGCTCTTTCGAAAGCC 653
    |||
200 rOLeuValCysGlyAspHisLeuArgGlyLeuValSerTrpGlyAsnIle 216
    |||
654 CCGTGCGCCAAAGTGGCGTGCGAGTGCTACACCAACCTCGCAAT 703
    |||
217 ProCysGlySerLysGluLysProGlyValTyrThrAsnValCysArgTy 233
    |||
704 CACTGAGTGATAGAGAAACCGTCAGGCC 734
    :|||
233 rThrasnTrpIleGlnLysThrIleGlnAla 243
```

**This Page Blank (uspto)**



OM of: US-09-030-606-173 to: SPTREMBL\_10.\* out\_format : pfs  
 Date: Sep 25, 1999 10:34 AM

About: Results were produced by the Gencore software, version 4.5,  
 Copyright (c) 1993-1998 Compugen Ltd.

## Command line parameters:

-MODEL-frame+ndp.model -DEV-xlp  
 -O-cgpn2.1/USPTO.spool/US09030606/runet.24091999.171617.29869/app\_query.fasta.1  
 -DB-SPTREMBL\_10 -OEMT-fastan -SUFFIX-rspt -GAPOP-12.000  
 -GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000  
 -GAPOP-4.500 -GAPEXT-0.050 -XGAPOP-10.000 -XGAPEXT-0.500  
 -FEAPOP-6.000 -FEAPEXT-7.000 -YGAPOP-10.000 -YGAPEXT-0.500  
 -DELOP-6.000 -DELEXT-7.000 -START-1 -MATRIX-biosum62  
 -TRANS-human0.cdi -LIST-45 -DOCALLIGN-200 -THR SCORE-escore  
 -ALIGN-15 -MODE-LOCAL -OUTFMT-pfs -NORM-stat -USER-US09030606  
 -NCPU-6 -ICPU-3 -WAIT -THREADS-1

## Search information block:

Query: US-09-030-606-173  
 Query length: 1265  
 Database: SPTREMBL\_10.\*  
 Database sequences: 201082  
 Database length: 61543640  
 Search time (sec): 260.540000

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
SP_human:0920M1	+	831.50	3.9e-64	255	0920M1 mus musculus (mouse). et
SP_human:075837	+	488.00	2.3e-34	282	0920M1 homo sapiens (human). se
SP_rhodent:063374	+	474.50	3.3e-33	235	063374 rattus norvegicus (rat). se
SP_rhodent:063375	+	468.00	1.2e-32	239	063375 rattus norvegicus (rat). se
SP_rhodent:092046	+	457.00	703.42	249	092046 dissotylus mawsoni. pr
SP_mammal:029474	+	456.50	702.23	261	029474 canis familiaris (dog).
SP_vertebrate:093365	+	452.00	695.31	250	093365 pleurocytes americanus.
SP_rhodent:088309	+	448.50	689.79	261	088309 mus musculus (mouse). pr
SP_rhodent:061855	+	436.00	670.34	261	061855 mus musculus (mouse). pr
SP_vertebrate:093366	+	433.50	667.11	242	093366 pleurocytes americanus.
SP_vertebrate:092099	+	432.50	665.56	242	092099 paraneothenia magellani.
SP_rhodent:054854	+	432.50	665.24	251	054854 rattus norvegicus (rat).
SP_vertebrate:042608	+	429.50	660.71	247	042608 petromyzon marinus (sea.
SP_vertebrate:042160	+	428.50	659.16	245	042160 petromyzon marinus (sea.
SP_rhodent:092189	+	423.50	651.45	245	092189 mus musculus (mouse). et
SP_vertebrate:042159	+	417.50	642.15	244	042159 petromyzon marinus (sea.
SP_rhodent:088301	+	414.00	636.96	237	088301 mus musculus (mouse). se
SP_vertebrate:091515	+	414.00	636.96	237	091515 fugu rubripes (Japanese.
SP_human:015665	+	407.50	626.08	259	015665 homo sapiens (human). et
SP_vertebrate:091501	+	392.50	603.63	234	091501 agkistrodon acutus (deli.
SP_rhodent:003955	+	388.50	596.83	250	003955 piramys natalensis (afr.
SP_vertebrate:013059	+	371.00	569.33	258	013059 trimeresurus gramineus (
SP_vertebrate:091516	+	359.00	550.67	258	091516 trimeresurus steinegeri.
SP_vertebrate:091505	+	359.00	549.98	279	091505 gloydius halys brevicaud
SP_vertebrate:091510	+	352.00	538.81	257	091510 gloydius ussuriensis. ca
SP_vertebrate:091516	+	352.00	539.71	260	091516 agkistrodon halys pallas
SP_vertebrate:091036	+	351.00	539.66	260	091036 gadus morhua (atlantic c
SP_vertebrate:091511	+	351.00	538.13	260	091511 gloydius halys (halys v4
SP_mammal:046683	+	349.50	536.13	251	046683 ovis aries (sheep). mast
SP_mammal:079343	+	348.50	531.01	377	079343 bos taurus (bovine). act
SP_vertebrate:091508	+	348.00	533.59	257	091508 trimeresurus mucroscuama
SP_vertebrate:091500	+	347.00	533.03	257	091500 trimeresurus mucroscuama
SP_vertebrate:091509	+	345.00	528.92	257	091509 trimeresurus mucroscuama
SP_vertebrate:091511	+	345.00	528.92	257	091511 trimeresurus mucroscuama
SP_vertebrate:091511	+	344.50	531.36	257	091511 trimeresurus mucroscuama
SP_vertebrate:091511	+	344.00	528.08	237	091511 agkistrodon halys pallas
SP_vertebrate:013060	+	340.00	522.04	260	013060 trimeresurus gramineus (
SP_vertebrate:013069	+	339.50	520.37	257	013069 bohyrops jarataca (jarat
SP_vertebrate:091519	+	339.00	519.59	257	091519 gloydius halys brevicaud
SP_mammal:029015	+	337.00	512.28	415	029015 sus scrofa (pig). prepic

SP\_vertebrate:091507 + 336.00 514.92 3.3e-21 257 | 091507 trimeresurus mucroscu  
 SP\_vertebrate:013057 + 335.00 513.27 4.1e-21 260 | 013057 trimeresurus flavovir  
 SP\_invertebrate:018599 + 334.50 512.62 4.5e-21 256 | 018599 drosophila virilis

seq\_name: SP\_rhodent:0920M1

seq\_documentation\_block:  
 ID 0920M1 PRELIMINARY; PRT; 255 AA.

AC 0920M1;  
 DT 01-MAY-1999 (TEMBREL. 10, Created)  
 DT 01-MAY-1999 (TEMBREL. 10, Last sequence update)  
 DT 01-MAY-1999 (TEMBREL. 10, Last annotation update)  
 DE EMBL MATRIX SERINE PROTEINASE 1 PRECURSOR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS-WEBSTER;  
 RA Slammer J.;  
 RT "Panel Matrix Serine Proteinase 1 (EMSP).";  
 RL Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF019979; AAC98894.1; -.  
 KW Signal.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 32 255 MATRIX SERINE PROTEINASE 1.  
 SQ SEQUENCE 255 AA; 27488 MW; 6E711616 CRC32.

## alignment\_scores:

Quality: 831.50 Length: 245  
 Ratio: 4.264 Gaps: 2  
 Percent Similarity: 79.592 Percent Identity: 62.857

## alignment\_block:

US-09-030-606-173 x 0920M1 ..

Align seg 1/1 to: 0920M1 from: 1 to: 255

4 AGCCCGGCACTGCGACGCTGCGAGCGGCGGCGACGTGATGAAAGCAATT 53  
 |||||  
 39 SerProHisSerGlnProTrpGlnAlaIleuPheSerGluAspGlyPh 55  
 |||||  
 54 GTTCGCTCGGGCGGCGGCGGCGATCGGCGGCGGCGGCGGCGGCGGCGAC 103  
 |||||  
 55 ePheCysSerGlyValIleValHisProGlnTrpValIleuSerAlaIle 72  
 |||||  
 104 ACTGTTCCAGAACCTCTACACACATCGGCGGCGGCGGCGGCGGCGGCGGAG 153  
 |||||  
 72 IScyLeuGlnIleuSerTrpIleValGlyLeuGlyLeuHisAsnIleuLys 88  
 |||||  
 154 GCCGACCAAGAACCGAGGAGCGGAGATGTTGGAGGCCAGCTTCGCGTAG 203  
 |||||  
 89 GlySerGlnGlnProGlySerArgMetLeuGlnAlaHisLeuSerIleG 105  
 |||||  
 204 GCACCAAGATGACACAGACCTTGCTGCTACAGCACTCATGCTCATCA 253  
 |||||  
 105 nHisProAsnPheAsnAspProSerPheAlaAsnAspLeuIleuLeu 122  
 |||||  
 254 AGTTGAGCAATCCGTGTCGAGTGCATGACACATCGGAGCATCGCAATT 303  
 |||||  
 122 YLeuAsnGlnSerValIleGlnSerAsnThrIleArgSerIleProVal 138  
 |||||  
 304 GCTTCGACATGCTTACCGCGGCGGAGAACTTGCCTGTTGCTGGG 353  
 |||||  
 139 AlaThrGlnCysProThrProGlySerPheCysLeuValSerGlyTyrG 155  
 |||||  
 354 TCTGTGCGCAAGCGGTGAGTCACGCGGTGTGTGCTGCCCTTCAAGA 403  
 |||||  
 155 YGlnIleuLysAsnGlyLys..... 161  
 404 GGTCTCTGCCACATGCGGCGGCGGATGACCAAGAGCTGCTGCCAGCA 453



Ratio: 2.893 Gaps: 7  
Percent Similarity: 64.822 Percent Identity: 40.316

Alignment block:  
US-09-030-606-173 x 063274 ..

Align seg 1/1 to: 063274 from: 1 to: 235

```

10 CACTGGACCCCTGGCAGCGGCGACTGGTCATGAAACGAATGTTCTG 59
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
10 AsnSerGlnProTyrGlnValAlaValAla.....AsnArgTyrLeu 24
60 CTCGGCGCTCTGTGTGTCATCCGACATGGGTGCTGTACGCCGACACTGT 109
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
24 sGlylyValLeuIleAspProSerTyrValIleThrAlaIleHisCysT 41
110 TCACAGACTCTACACACATCGGCGCTGGCGCTGCACAGTCTTGAGCCGAC 159
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
41 ySerHisTyrTyrHisValLeuLeuGlyArgHisAsnLeuHeGluAsp 57
160 CAAGACCCAGGAGCGAGATGGTGGAGCGACCTCCGCTACGGCAGCC 209
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
58 ..GluProPheAlaGlnTyrArgPheValSerGlnSerPheProHisPr 73
210 AGAGTACACACAGACCCCTGCTC..... 231
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
73 OAspTyrAsnProPheLeuMetArgAsnHisThrArgGlnThrGlyTyrA 90
232 ....GCTAACGACCTCATGCTCATCAAGTGGAGCAATCCGTCGCCAG 276
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
90 sPtyrSerAsnAspLeuMetLeuHisLeuSerGluProAlaAspIle 106
277 TCTGACACATCCGGACGATGAGCATGCTGTGGAGTCCGTCACGCGGG 326
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
107 ThrAspGlyValIysValIleAspLeuProThrGlnGluProIysValGl 123
337 GAACCTCTTCCTGTTTCTGAGTGGGTCTGCTGGCAACGGTGAGCTCA 376
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
123 ySerThrCysLeuValSerGlyTyrPly..... 132
377 CGGGTGTGTGCTGCCCTTTCAGAGAGGCTCTGCCCCAGTCGGCGGGG 426
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
133 .....SerThrLys..ProLeuIleSer..... 139
427 CTGACCCAGAGCTGTGCTGCCAGGACGAATGCCCTACGCTGCTGACATGC 476
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
140 .....Glu..LeuProAspAspLeuGlnCys 147
477 GTGAACGTGTGCGTGTCTGAGAGAGTCTGACAGTAACTCTATGACCC 526
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
148 ValAsnIleAspLeuLeuSerAsnGluYscYsIleGluAlaIleTyrArgTr 164
527 GCTGTACACACCCAGCATGTTCTGCGCGCGCGGAGGCGACAGACCAAG 576
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
164 pLysValIThrAspLeuMetLeuCysAlaIleIyLysLeuGlnGlyLysA 181
577 ACTCTGACAGGTGACTCTGGGGGCGCCCTGATGCTGCAACGGGTAATCG 626
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
181 sPalCysAsnGlnIysAspSerGlyGlyProLeuIleCysAspIlyValLeu 197
627 CAGGCGCTTGTGCTTTCGAAAAGCCCGCTGTGCGCAAGTGGCGTCC 676
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
198 GlnGlyLeuThrSerTyrPlySerValProCysGlyGluProHisAsnPr 214
677 AGGTGCTTACACCACTCTGCAAAATTCAGTAGAGGATAGAAAACCG 726
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
214 oGlyIleTyrThrIlySleIleIlySpherHisSerTyrIleLysGluVal 231
727 TCCAG 731
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
231 eLys 232

```

seq\_name: sp\_rdent:063275

seq\_documentation\_block:

```

ID 063275; PRELIMINARY; PRT; 239 AA.
AC 063275;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, last annotation update)
DE KALLIKREIN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
RA ZINTZ C.B., MA J.X., CHAO J., CHAO L.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; L33840; AA58782.1; -.
DR PIRAM; PF00089; trypsin; 1.
FT NON_TER
FT SEQUENCE 1 239 AA; 26382 MW; 0609E3E6 CRC32;

```

alignment\_scores:

Quality: 469.00 Length: 252  
Ratio: 2.825 Gaps: 6  
Percent Similarity: 65.873 Percent Identity: 39.286

Alignment block:

US-09-030-606-173 x 063275 ..

Align seg 1/1 to: 063275 from: 1 to: 239

```

10 CACTGGACCCCTGGCAGCGGCGACTGGTCATGAAACGAATGTTCTG 59
   ::::::::::::::::::::::::::::::::::::::::::::::::::::
18 AsnSerGlnProTyrGlnValAlaValAla.....sGlnAspLeu 32
60 CTCGGCGCTCTGTGTGTCATCCGACATGGGTGCTGTACGCCGACACTGT 109
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
32 sGlylyValLeuIleAspProSerTyrValIleThrAlaIleHisCysT 49
110 TCACAGACTCTACACACATCGGCGCTGGCGCTGCACAGTCTTGAGCCGAC 159
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
49 ySerAspAsnTyrHisValLeuLeuGlnAsnAsnLeuSerGluAsp 65
160 CAAGACCCAGGAGCGAGATGGTGGAGCGACCTCTGCTGACGACCC 209
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
66 ValGln.....HisArgLeuVal.....SerGlnSerPheArgHisPr 78
210 AGAGTACACACAGACCCCTGCTC..... 231
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
78 OAspTyr..LysProPheLeuMetArgAsnHisThrArgIysProIysA 94
232 ....GCTAACGACCTCATGCTCATCAAGTGGAGCAATCCGTCGCCAG 276
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
94 sPtyrSerAsnAspLeuMetLeuHisLeuSerGluProAlaAspIle 110
277 TCTGACACATCCGGACGATGAGCATGCTGTGGAGTCCGTCACGCGGG 326
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
111 ThrAspGlyValIysValIleAspLeuProThrIlySglnProIysValGl 127
337 GAACCTCTTCCTGTTTCTGAGTGGGTCTGCTGGCAACGGTGAGCTCA 376
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
127 ySerThrCysLeuValSerGlyTyrPly..... 136
377 CGGGTGTGTGCTGCCCTTTCAGAGAGTCTGCCCCAGTCGGCGGGG 426
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
136 ..... 136
427 CTGACCCAGAGCTGTGCTGCCAGGACGAATG..CCTACCGTGTGACTG 475
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
137 .....SerThrAsnProSerGlnTyrPlyGluPheProAspAspLeuGln 151
476 CGTGAACGTGTGCGTGTCTGAGGAGGCTGCGAGTCAAGCTCATATGACC 525

```



```

10 CACTGGCAACCCCTGGCAGGCGGACGGTCATGGAACAAACGATGTTCTG 59
11 |||||
32 HISSERAlgProTYMeAlaSerLeuAsnTYrGlyYrHis..PheCY 47
33 |||||
60 CTGGGCGCTCTGTGTCATCCGCAAGTGGTGCTCTCAAGCCGACACTGT 109
61 |||||
47 SelGYAlaLeuIleAsnAsnGlnTPrValLeuSerValAlaHisCysT 64
48 |||||
110 TCACAGACTCTTACACC.....ATCGGCGCTGGGCGCTGCACAGTCTTGAG 155
111 |||||
64 rPrTYAsnPrOTyrAlaMetClnValMetLeuGlyGlnHisAsnLeuArg 80
65 |||||
154 GCGGCAACAAAGCCAGGGAGCCAGATGGTGTGAGGCCAGCGCTCTCCGACG 203
155 |||||
81 Val...PheGlnGlyThrGlnIleuMetCysThrAspHrIleLeuT 96

```

```

204 GCACCCAGAGTACACAGACCCCTTGCTGCTAACGACCTCATGCTCATCA 253
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
96 pHisProAsnTyrAspTyrGlnThrLeuAspPheAspLeuMetLeuIleL 113
254 AGTTGACAGAAATCCGTGCGAGTCTGACACCAATCCGGACATACAGCAT 303
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
113 ySLeUTyrHisProValGluValThrGlnAlaValAlaProIleSerLeu 129
304 GCTTCGAGAGTCCCTACCGGGGGGAGACTCTGCTCGTTTGGCTGGGG 353
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
130 ProSerSerCysProValGlyGlyThrProCysSerValSerGlyTyrPgl 146
354 TCTCTGCGCAGAACGCTGAGCTCAGCGGTGTGTCTGCTCCCTTCAAGGA 403
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
146 y..... 146
404 GGTCTCTGCCCACTCGCGGGGCTGACCCAGACGCTGCTGCCAGGCA 453
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
147 .AsnThrAlaArgAsp..GlyAspAspVal.....T 156
454 GAATGCCCTACCGTGTGAGTGGTGAACGTTGGTGGTGTGCTGAGGAG 503
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
156 yrmcProThrLeuLeuGlnCysMetAspValProIleIleAspGlu 172
504 GTCTGACGTAAGCTCTATGACCCGCTGACCAACCCAGCATGTTCTGCGC 553
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
173 GlnCysMetLysSerTyrProGlyMetLeuSerProArgMetValCysAl 189
554 CGCGCGAGGCGACAGACCAAGAGACTCTGCAACGGTGACTCTGGGGGC 603
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
189 aGlyPheMetAspGlySerArgAspAlaCysAspGlyAspSerGlySerP 206
604 CCCGATCTGCAACGGGACTTGCAGAGGCGCTGTGTCTTGGGAAAGCC 653
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
206 rGluValCysArgGlyGluValThrGlyLeuValSerTyrPglGlyMet 222
654 CCGTGTGGCCAAAGTTGGCGTGCAGAGTGTCTACACCACTCTCAAAAT 703
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
223 ..CysAlaGlnProAsnTyrProGlyValGlyValLysLeuGlyCysGluPh 238
704 CACTGAGTGGATAGAGAAACCTCCAGGCC 734
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
238 eHisAlaTyrPglGlnAsnThrLeuAlaAla 248

seq_name: sp_rJodent:088309
seq_documentation_block:
ID 088309 PRELIMINARY: PRT: 261 AA.
AC 088309;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE PRORENIN-CONVERTING ENZYME (MK13B) PRECURSOR (MK13B).
GN MKL-13B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98351995.
RA HOSOI K., TADA J., TSUMURA K., KANAMORI N., YAMANAKA N.;
RT "Expression of an allozyme of prorenin-converting enzyme in the
RT submandibular gland of DBA/2N mice.";
RL J. Biochem. 124:368-376(1998).
DR EMBL; AB016032; BAA3186.1; -.
DR PFM; PF00089; trypsin; 1.
KW Signal.
FT SIGNAL 1 24 POTENTIAL.
SO SEQUENCE 261 AA; 28692 MW; 8DB1814F CRC32;

```

alignment\_scores: Quality: 448.50 Length: 252

```

Ratio: 2.786 Gaps: 6
Percent Similarity: 63.889 Percent Identity: 38.492
alignment_block:
us-09-030-606-173 x 088309 ..
Align seg 1/1 to: 088309 from: 1 to: 261

10 CACTCGAGGCCCTGGCAGCGGGCAGCTGTCATGAAAGAAATGTTCTG 59
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
34 AsnSerGlnProThrGlnValAlaValAlaValTyrTyrGlnLysGlnHisTleCy 50
60 CTGGGCGCTCTGTCATCCGAGTGGGTGCTGTGACGCGCACATGTT 109
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
50 sGlyGlyValLeuLeuAspArgAsnTyrPglValLeuThrAlaAlaHisCysT 67
110 TCCAGAACTCTTACACCATGGGCTGGGCTGCACAGTCTTGAGGCCGAC 159
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
67 yValAspGlnTyrGluValTyrPglGlyLysAsnLysLeu...PheGln 82
160 CAGAGCCAGGAGCCAGATGTTGAGGCCAGCCCTCCGTCACGACCC 209
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
83 GlnGluProSerAlaGlnHisArgLeuValSerLysSerPheProHisP 99
210 AGAGTACAAAGACCCCTTGCTC..... 231
99 oGlyTyrAsnMetSerLeuLeuMetLeuGlnThrProGlyAlaAla 116
232 ....GCTAACGACTTCATGCTCATCAAGTTGAGCAAAATCGTGTCCGAG 276
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
116 sPheSerAsnAspLeuMetLeuArgLeuSerLysProAlaAspIle 132
277 TCTGACACATCCGAGCATACAGCATGTTGCTGCAATGCCCTACCGCGGG 326
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
133 ThrAspAlaValLysProIleAlaLeuProThrLysGlnProLysProG 149
327 GAACCTTGCCCTGTTTCTGCTGGGTGCTGTCGCGCAACGCTGACTCA 376
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
149 ySerLysCysLeuAlaSerGlyTyrPglYserIle..... 160
377 CGGGTGTGTGTGCTCCCTTCAAGAGGCTCTGTCCCAAGTCCGCGGGG 426
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
161 .....ThrProThrArgTyr.. 165
427 CTGACCCAGAGCTGTGGTCCAGGAGCAATGCCCTACCGTGCAGTGC 476
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
166 .....GlnLys.ProAspAspLeuGlnCys 173
477 GTGAACTGTGCGGTGTGTCTGAGAGGTCGTCAGTAAGCTATGACCC 526
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
174 ValPheLeuThrLeuLeuProAsnGlnAsnCysAlaLysValTyrLeuG 190
527 GCTGTACACCCAGCATGTTGTGCGCGCGGAGGAGGCAACAGCAAGG 576
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
190 nLysValThrAspValMetLeuCysAlaGlyGluMetGlyGlyLysA 207
577 ACNCCCTCAACGCTGACTCGGGGGGCCCTGATCTGCACAGGCTACTTG 626
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
207 sPthrCysArgAspAspSerGlyGlyProLeuIleCysAspGlyIleLeu 223
627 CAGGCGCTGTGCTTCCGAAAGCCCGTGGCCCAAGTTGCGCTGCG 676
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
224 GlnGlyThrThrSerTyrGlyProThrProGlyLysProGlyValP 240
677 AGGTGCTACCAACCTTGCAAAATTCAGTGAAGGATAGAGAAACCG 726
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
240 oAlaIleTyrThrAsnLeuIleLysPheAsnSerTyrIleLysAspThrM 257
727 TC 728
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
257 et 257

seq_name: sp_rJodent:061855

```

seq\_documentation\_block:  
 ID 061855 PRELIMINARY; PRT: 261 AA.  
 AC 061855;  
 DT 01-NOV-1996 (TREMblrel. 01, Created)  
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
 DT 01-JAN-1999 (TREMblrel. 09, Last annotation update)  
 DE TISSUE KALLIKREIN.  
 GN MGK-6.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA TADA M., PETERS J., TAKAHASHI S., INOUE H., MIYAKE Y.;  
 RT "Identification of a tissue kallikrein gene, mgk-6, expressed in a  
 RT mouse neuroendocrine cell line."  
 RL Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; D10464; BAA01257.1; -  
 DR PFM; PF00089; trypsin; 1.  
 SO SEQUENCE 261 AA; 28775 MW; 509D1D00 CRC32;

alignment\_scores:  
 Quality: 436.00 Length: 251  
 Ratio: 2.691 Gaps: 4  
 Percent Similarity: 64.542 Percent Identity: 36.255

alignment\_block:  
 US-09-030-606-173 x 061855 ..

Align seg 1/1 to: 061855 from: 1 to: 261

```

10 CACATCGCAGCCCTGGCAGGCGGCGGCTGCATGCAAAACGAATTGTTCTG 59
   |||||||
34 ASnergtglnprptglnvalalavalayrgrphetrlytyglnscy 50
   |||||||
60 CTCGGCGCTCTGGTGCATCGGAGTGGTCTGTCAGCCGACACATCTT 109
   |||||||
50 SGtGtYtLeuLeuasnAlaAsnTPValLeuThraAlaAlaScysH 67
   |||||||
110 TCAGAACCTCTACACATCGGCGTGGCTGTCAGTCTTGAGGCGGAC 159
   |||||||
67 lAsnAspLytyglnvaltgrpleuglylAsnAsnRheleuglnsyr 83
   |||||||
160 CAGAGCCGAGGAGCAGATGTTGAGGCGGACCTCTCGTACGGCAGCC 209
   |||||||
84 ..gluprosertAlaGlnAlaArgLeuValSerLySaAlaIleProAlaPr 99
   |||||||
210 AGAGTACAAGACCCCTGCTC..... 231
   |||||||
99 oAspRheAsnMetSerLeuLeuAsnGlnAlaThrProGlnProGlnAsp 116
   |||||||
232 .....GCTAACGACCTCATGCTCATCAAGTGAAGATCCGTCGAG 276
   |||||||
116 spYrSerAsnAspLeuMetLeuLeuArgLeuLylylProAlaAspRle 132
   |||||||
277 TCTGACACCATCGGAGCATGAGCATGCTTGGAGTCCCTTACCGCGGG 326
   |||||||
133 ThrAspValVallylProIleAspLeuProThrglnGlnProLyLeuGln 149
   |||||||
327 GAGCTCTGCTGCTGCTGGTGGGCTGCTGCGGAAGCGTGAAGTCA 376
   |||||||
149 ySerThrgyLeuAlaSerGlyTTP.....GlySerIleT 161
   |||||||
377 CGGtGtGCTCTGCTGCTCTTCAAGAGAGCTCTGCCAGTCCGGGCGG 426
   |||
161 hrProVal..... 163
   |||
427 CTGACCCGAGACTGCTGCCAGGCGAGATG..CTTACCGTGGTCACTG 475
   |||
164 .....LysTygIleTyTTProAspGlnLeuGlnCy 173

```

```

476 CGTGAACGTGTCGTGCTGTGAGAGAGTCTGCAGTAACTATGACC 525
   |||||||
173 sValAsnLeuLySerLeuProAsnGlnAspCyAlaLyAlaAlaHisIleG 190
   |||||||
526 CGCTGTACACCCGACATGTTTGGCCGCGGAGGCGCAAGACAGAG 575
   |||||||
190 lulyValThrAspAspMetLeuCySalaglyAspMetAspGlyLylys 206
   |||||||
576 GACTCTCGCAACGGTGACTCGGGGCGCCCTGTGTCAGCAGGCTACT 625
   |||||||
207 AspThrcySalaglyAspSerGlyGlyProLeuIleCyAspGlyValLe 223
   |||||||
626 CAGAGCGCTTGTCTTTCGGAAGCCCGTGGCCAGTGGCGCTGC 675
   |||||||
223 uGlnGlyIleTherSerTrpGlyProSerProCyGlyLyAsProAsnValP 240
   |||||||
676 CAGGTGTCTACACCACTCTGCAATTCACGTAGTGAAGAGAAACC 725
   |||||||
240 rogIlyIleTyThrArgValLeuAsnRheAsnThrTrpIleArgGlnThr 256
   |||
726 GTC 728
   |||
257 Met 257

```

seq\_name: sp\_vertibrate:093266

seq\_documentation\_block:  
 ID 093266 PRELIMINARY; PRT: 242 AA.  
 AC 093266;  
 DT 01-NOV-1998 (TREMblrel. 08, Created)  
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)  
 DT 01-MAY-1999 (TREMblrel. 10, Last annotation update)  
 DE TRYPSINOGEN 2 PRECURSOR (EC 3.4.21.4).  
 GN TRP2.  
 OS Pleuronectes americanus.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;  
 OC Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Percomorpha;  
 OC Pleuronectiformes; Pleuronectoidae; Pleuronectidae; Pleuronectes.  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-INTESTINE, PYLORIC CAECA, AND PANCREAS;  
 RA DOUGLAS S.E., GALLARD J.W.;  
 RT "Isolation of cDNAs for Trypsinogen from the Winter Flounder,  
 RT Pleuronectes americanus."  
 RL J. Mar. Biotechnol. 0:0-0(1998).  
 DR EMBL; AF012463; AAC32752.1; -  
 DR PFM; PF00089; trypsin; 1.  
 KW Signal; Hydrolase.  
 FT SIGNAL 1  
 FT CHAIN 21  
 FT SIGNAL 15  
 FT CHAIN 21  
 FT SIGNAL 15  
 SO SEQUENCE 242 AA; 26180 MW; 7E7ZC91C CRC32;

alignment\_scores:  
 Quality: 433.50 Length: 242  
 Ratio: 2.660 Gaps: 6  
 Percent Similarity: 67.355 Percent Identity: 38.017

alignment\_block:  
 US-09-030-606-173 x 093266 ..

Align seg 1/1 to: 093266 from: 1 to: 242

```

4 AGCCGCACTGCGACGCTGCGAGCGGCGGCACTGTCATGGAACGAATT 53
   |||||||
28 ThrProHisSerGlnAlaHisGlnValSerLeu...AsnSerGlyTyrH 43
   |||||||
54 GTTCTGCTGGGCTCTGTCGCAATCCGAGTGGTCTGTCACCCGAC 103
   |||||||
43 sPhcCyGlyGlySerLeuValaGlnAsnTPValValSerIleAlaH 60
   |||||||
104 ACTTTTCAGAACTCTTACACATCGGCGTGGGCTGTCAGAGTGTGAG 153
   |||||||

```





158 ..... 158  
231 easnsptrleugltnrsermet 239

seq\_name: sp\_rodent:054854

seq\_documentation\_block: PRT: 251 AA.  
ID\_054854 PRELIMINARY;  
AC 054854;  
DT 01-JUN-1998 (TREMBLrel. 06, Last Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-MAY-1998 (TREMBLrel. 10, Last annotation update)  
DE MYELENCEPHALON SPECIFIC PROTEASE.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY;  
RX MEDLINE: 97477435.  
RA SCARISBRICK I.A., TOWNER M.D., ISACKSON P.J.;  
RT "Nervous system-specific expression of a novel serine protease;  
RT regulation in the adult rat spinal cord by excitotoxic injury";  
RL J. Neurosci. 17:8156-8168(1997).  
DR EMBL: AF016269; AAC02300.1; -  
DR PFM: PF00089; trypsin.1.  
KW Protease.  
SQ SEQUENCE 251 AA; 28014 MW; 21D0D8A9 CRC32;

alignment\_scores:  
Quality: 432.50 Length: 247  
Ratio: 2.755 Gaps: 4  
Percent Similarity: 63.563 Percent Identity: 36.032

alignment\_block:

us-09-030-606-173 x 054854 ..

Align seg 1/1 to: 054854 from: 1 to: 251

1 GGCAGCCG.....CACTGCGACCCCTGGACGGCGCACTGTCAT 41  
|||||  
32 GlytGlyProCysLeuLysAsnSerHisProPheGlnAlaLeuIyrth 48  
42 GGAAGAAGCATGTTCTGCTCGGCGCTCTGTCATCGCGCATCGGTC 91  
|||||  
48 rSerGlyHisLeuLysGlyGlyValLeuValGlyProGlnIyrth 65  
92 TGTCAAGCGCACACTGTTCCAGAACTCTACACCATCGGCTGGCCCTG 141  
|||||  
65 eutThrAlaLeuHisCysLysLysProAsnLeuGlnValIyrLeuGlyLys 81  
142 CACACTCTTGAAGCGGACCAAGCGGACCGACAGAGTGGAGCCAG 191  
|||||  
82 HisAsnLeuArg...GlnThrGluThrPheGlnArgGlnIleSerValas 97  
192 COTCCGCGACGCGACAGAGTACAGACACCTTGCCTGCGTACAGACC 241  
:|||||  
97 PArgThrIleValHisProArgTyrAsnProGlnThrHisAsnAspI 114  
242 TCATCTCTCATCAAGTTGAGCAATCCGTGCGAGTCTGACACCATCGG 291  
:|||||  
114 leuMetValHisLeuLysArgProValLysPheSerGlnArgIleGln 130  
292 AGCATCAGCATGCTTGCAGTGCCTTACCGCGGGAACCTTGCCTCGT 341  
:|||||  
131 ProLeuProLeuLysLysAspCysSerGlnLysAsnProAspCysGlnI 147  
342 TTCCTGCTGGGGTGTGCTGCGCAAGCTAGCTACAGGGTGTGTCTCTC 391  
|||||  
147 eueuGlyTyrGlyLysMetGlnAsnGlyu...Phe..... 158  
392 CCTCTTCAAGAGAGTCTCTGCGCCAGTCCGGGGGCTGACCCAGACTCT 441

158 ..... 158  
442 GCGTCCAGCAGAAATGCGTACCGTCTGACGTGGTGAAGCTGCGTG 491  
|||  
159 .....ProAspThrIleGlnCysAlaAspAlaGlnLeu 169  
492 GTGTCTGAGAGTCTGCGATGAGCTGTATGACCGCGTGTACCAACCCAG 541  
|||||  
170 ValSerAlaGluGluCysGluArgAlaIyrProGlyLysIleThrArgSe 186  
542 CATGTTCTGCGCGCGCGGAGGCAAGACCAAGAGACTCTGCAACGGTG 591  
|||  
186 rMetValCysAlaGlyAspLysArgGluGlyLysAsnAspCysGlnGly 203  
592 ACTCTGGGGGGCCCTGATCTGCAACGGGTACTGTCCAGGCGCTGTCT 641  
|||||  
203 spSerGlyGlyProLeuValCysGlyGlyHisLeuValGlyIleValSer 219  
642 TTCGGAAGACCCCGTGTGCGCAAGTGGCGCCAGGTGTCTACCAACA 691  
:|||||  
220 TrpGlyAspMetProCysGlySerLysGluLysProGlyValIyrThrAs 236  
692 CCTCTGCAAAATTCAGTGAAGTGAAGAAACCGTCCAG 731  
:|||||  
236 pValCysThrHisIleArgTyrIleGlnAsnIleIleArg 249

seq\_name: sp\_vertebrate:042608

seq\_documentation\_block: PRT: 247 AA.  
ID\_042608 PRELIMINARY;  
AC 042608;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
DE TRYPsinogen A3 PRECURSOR.  
GN TRYP3.  
OS Petromyzon marinus (Sea lamprey).  
OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi;  
OC Petromyzontiformes; Petromyzontidae; Petromyzon.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA ROACH J.C.;  
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=ANTERIOR INTESTINE;  
RA ROACH J.C.;  
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF011899; AAB69655.1; -  
DR EMBL: AF011352; AAB65411.1; -  
DR PFM: PF00089; trypsin.1.  
KW Signal.  
FT SIGNAL 1 15 POTENTIAL.  
FT CHAIN 16 247 TRYPsin A3.  
SQ SEQUENCE 247 AA; 26295 MW; 6D71AC2E CRC32;

alignment\_scores:  
Quality: 429.50 Length: 245  
Ratio: 2.635 Gaps: 6  
Percent Similarity: 66.531 Percent Identity: 36.327

alignment\_block:

us-09-030-606-173 x 042608 ..

Align seg 1/1 to: 042608 from: 1 to: 247

10 CACTGCGACCCCTGCGACGCGCACTGTCATGGAAGAATTTCTG 59  
|||||  
33 HisSerGlnProTyrGlnValSerLeuAsnIleGlyTyrHis...Phe 48  
:|||||  
60 CTCGGGCTCTGTCATCCGCAAGTGGTGTCTGACCGCCACACTGTT 109  
:|||||

48 sgllyglSerLeuLeuIleasnSerGIIntrPrValValSerAlaAlaHisCysT 65

110 TCAGAACTCC.....TACACATCGGGCTGGGCTGCACAGTCTTGAG 153

65 yrlIntrIntraSerArgIleSerValArgIleGIyLInHisnIlePhe 81

154 GCGGACCAAGAGCCAGGGAGGCAGATGGTGGAGGCGCAGGCTCTCGTAG 203

82 ValAsn...GIuLyThrGIuLGIleGIInGlnAlaSerLysAlaIleGI 97

204 GCACCCAGAGTACAAACAGACCCCTTGCTGCTACACGACCACTCATCA 253

97 nHISPrOGInTyrAsnSerTrpThrIleAspAsnSpmIleMetLeuIle 114

254 AGTTGACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCAAT 303

114 yslSerSerProAlaThrLeuAsnGlnTyrAlaGlnAlaIleAlaLeu 130

304 GCTTCGAGTGCCTTACCGGGGAACTTGGCTTCCTTTCGGCTGGGG 353

131 ProSerSerLysValAsnThrGIyAlaMetCysThrIleSerGIyTrpGI 147

354 TCTGTGGCGAACGCTGAGCTCACGGGTGTGTGTGCTCCCTTTCAAGA 403

147 Y..... 147

404 GGTCTCTGCCACTCGCGGGGCTGACCCAGAGCTGTGCTCCAGCA 453

148 .....GIuThrGIInThrSerValGIy 154

454 GAATGCTACCGTCTGTCAGTGCAGACGTGTGCTGTGTGTAGGAG 503

155 Ser..ProAspValLeuMetCysValGlnAlaProValLeuSerAspThr 170

504 GTCTGCAGTAACTCTATGACCCCGCTGTACACCCAGCATTTCTGGC 553

171 SerCysArgAsnSerTyrProGIyAspIleThrAsnAsnMetIleCysIe 187

554 CGCGCGAGGCGAAGCAGACGAAGACTCCTGCACAGGTGACTGTGGGGGC 603

187 uGIyTyrLeuGIuLGIyGIyAspSerCysGlnGIyAspSerGIyGIy 204

604 CCTGATCTGCACAGGGGACTTGCAGGGGCTGTGTCTTTCGGAAGCC 653

204 rovalValCysAsnGIyGIuLeuGlnGIyIleValSerTrpGIyAsnGIy 220

654 CCGTGTGGCAAGTGTGGCGGTGCTGTCTACACCAACCTTCGAAT 703

221 ..CysAlaLeuProAsnTyrProGIyAlaTyrThrLysValCysAsnGIy 236

704 CACTGACTGGATAGAAACCGTCCAGCGCAGT 737

236 rAsnAlaTrpIleAlaGlnThrIleAlaAlaAsn 247

seq\_name: sp\_vertibrate:042158

seq\_documentation\_block:

ID 042158 PRELIMINARY; PRT; 247 AA.

AC 042158;

DT 01-JAN-1998 (TREMblrel. 05, Created)

DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)

DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)

DE TRYPsinOGEN A2 PRECURSOR.

GN TRYP2.

OS Petromyzon marinus (Sea lamprey).

OC Eukaryota; Metazoa; Chordata; Craniata; Cephalaspidomorphi

NC Petromyzontiformes; Petromyzontidae; Petromyzon.

RA ROACH J.C.;

BL Submitted (Jul-1997) to the EMBL/GenBank/DBJ databases.

EMBL: AF011898; AAB69654.1; -.

PIR: PF00089; trypsin: 1.

DR

[illegible]

Db 330 TGGTTTGATCATCTCTCTTTTCCCATTCGAACTAGTCATTAAACCATCTCTGAA 271  
OY 182 CTGTAGAAAAACATCTGAAGAGCTAGTCTATCGACATCGACAGTGAAATGGATGTT 241  
Db 270 CTGTAGAAAAACATCTGAAGAGCTAGTCTATCGACATCGACAGTGAAATGGATGTT 211  
OY 242 CTCGAAACATTTTACCACAGACACCTGTTTCTATCCTGTTTAAATAATTAGTTGGTT 301  
Db 210 CTCGAAACATTTTACCACAGACACCTGTTTCTATCCTGTTTAAATAATTAGTTGGTT 151  
OY 302 CTCACATGATTAACAACACCTGCTCCAAATCTGTACATAAAAGCTGTGACTGGAAGTT 361  
Db 150 CTCACATGATTAACAACACCTGCTCCAAATCTGTACATAAAAGCTGTGACTGGAAGTT 91  
OY 362 TAGTC 366  
Db 90 TAGTC 86

RESULT 3  
LOCUS AA809587 435 bp mRNA EST 18-FEB-1998  
DEFINITION nJ43611.s1 NCI-CGAP\_P9 Homo sapiens cDNA clone IMAGE:595276, mRNA  
ACCESSION AA809587  
NID 92878993  
VERSION AA809587.1 GI:2878993  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 435)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jan 19, 1998 this sequence version replaced gi:2287384.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuagui,  
M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbip/image/image.html

Insert Length: 605 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham.  
Location/Qualifiers  
1. 435  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="902E09; 2: 2p11.2-2p13.1"  
/clone\_image="IMAGE:595276"  
/clone\_lib="NCI-CGAP\_P9"  
/sex="male"  
/tissue\_type="normal prostatic epithelial cells"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: PAMP10; mRNA made from  
normal prostatic epithelial cells, cDNA made by oligo-dT  
priming. Non-directionally cloned. Size selected on  
agarose gel, average insert size 600 bp. Library made by  
D. Kitzman, NIH"

BASE COUNT 120 a 96 c 68 g 151 t  
ORIGIN

Query Match 99.3%; Score 363.4; DB 38; Length 435;

Best Local Similarity 99.7%; Pred. No. 2e-73;  
Matches 364; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CTCTTTCTCCCTCCTCTGAAATTAATCTTTCAACTTGCATTTGCAAGATTACAC 61  
Db 21 CTCTTTCTCCCTCCTCCTCTGAAATTAATCTTTCAACTTGCATTTGCAAGATTACAC 80  
OY 62 ATTTCACTGATGATTAATGTTGTGCAAAAAAAGAGCTGTTTAAATAATTA 121  
Db 81 ATTTCACTGATGATTAATGTTGTGCAAAAAAAGAGCTGTTTAAATAATTA 140  
OY 122 TGGTTTGATCATCTCTCTTTTCCCATTCGAACTAGTCATTAAACCATCTCTGAA 181  
Db 141 TGGTTTGATCATCTCTCTTTTCCCATTCGAACTAGTCATTAAACCATCTCTGAA 200  
OY 182 CTGTAGAAAAACATCTGAAGAGCTAGTCTATCGACATCGACAGTGAAATGGATGTT 241  
Db 201 CTGTAGAAAAACATCTGAAGAGCTAGTCTATCGACATCGACAGTGAAATGGATGTT 260  
OY 242 CTCGAAACATTTTACCACAGACACCTGTTTCTATCCTGTTTAAATAATTAGTTGGTT 301  
Db 261 CTCGAAACATTTTACCACAGACACCTGTTTCTATCCTGTTTAAATAATTAGTTGGTT 320  
OY 302 CTCACATGATTAACAACACCTGCTCCAAATCTGTACATAAAAGCTGTGACTGGAAGTT 361  
Db 321 CTCACATGATTAACAACACCTGCTCCAAATCTGTACATAAAAGCTGTGACTGGAAGTT 380  
OY 362 TAGTC 366  
Db 381 TAGTC 385

RESULT 4  
LOCUS AA813266/c 599 bp mRNA EST 31-DEC-1998  
DEFINITION aJ44603.s1 Soares\_testis\_NHT Homo sapiens cDNA clone 1393204 3',  
mRNA sequence.  
ACCESSION AA813266  
NID 92883251  
VERSION AA813266.1 GI:2883251  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 599)  
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLES National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Apr 18, 1995 this sequence version replaced gi:775316.

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima  
Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
www.bio.llnl.gov/dbip/image/image.html

Insert Length: 982 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 503.  
Location/Qualifiers  
1. 599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="1393204"  
/clone\_lib="Soares\_testis\_NHT"  
/sex="male"

FEATURES  
source

/lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - o1190(dT)  
 primer [5].  
 TGTACCAATCGAAGTGGAGCGGCCCAATTTTATTTTATTTT 3'}.  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
 ORIGIN

202 a 98 c 131 g 168 t

Query Match 96.7%; Score 354; DB 39; Length 599;  
 Best Local Similarity 99.7%; Pred. No. 2.6e-71;  
 Matches 365; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

2 CTTCTCTCCCTCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTAC 61  
 454 CTTCTCTCTCCCTCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTAC 395  
 62 ATTTCACGTGATGATATATGTTGC-AAAAAAAAAAAGTCTTTGTTAAATTAC 120  
 394 ATTTCACGTGATGATATATGTTGC-AAAAAAAAAAAGTCTTTGTTAAATTAC 335  
 121 TTGTTTGTGATGATGATGTTGTTTCCCATTTGCAATGATTAACCATCTCTGA 180  
 334 TTGTTTGTGATGATGATGTTTCCCATTTGCAATGATTAACCATCTCTGA 275  
 181 ACTGATGAAAAACATCTGAAAGCTAGTCTATCAGCATGTCAGGTAATGGATGT 240  
 274 ACTGATGAAAAACATCTGAAAGCTAGTCTATCAGCATGTCAGGTAATGGATGT 215  
 241 TCTCAGAACCATTTTCAACCCAGACAGCTGTTTCTATCCTGTTTAAATTAATTTGGGT 300  
 214 TCTCAGAACCATTTTCAACCCAGACAGCTGTTTCTATCCTGTTTAAATTAATTTGGGT 155  
 301 TCTCTACATGATTAACAACCCCTGCTGCAATGTCATCAATAAAGTGTGACTTGAAGT 360  
 154 TCTCTACATGATTAACAACCCCTGCTGCAATGTCATCAATAAAGTGTGACTTGAAGT 95  
 361 TTAGTC 366  
 94 TTAGTC 89

RESULT 5  
 AA437224/c 490 bp mRNA EST 30-MAY-1997  
 LOCUS zvs5b10.s1 Soares\_testis\_NHT Homo sapiens cDNA clone IMAGE:757435  
 DEFINITION 3' similar to contains Alu repetitive element; mRNA sequence.  
 ACCESSION AA437224  
 NID 92142138  
 VERSION AA437224.1 GI:2142138  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 490)  
 Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,  
 Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J.,  
 Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,  
 White, Y., Wylie, T., Waterston, R., and Wilson, R.  
 WashU-Merck EST Project 1997  
 Unpublished (1997)  
 On Sep 12, 1996 this sequence version replaced gi:1405293.

Contact: Wilson RK  
 Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
 Tel: 314 286 1800  
 Fax: 314 286 1810  
 Email: estevanston.wustl.edu  
 This clone is available royalty-free through LNL; contact the  
 IMAGE Consortium (info@image.llnl.gov) for further information.  
 Seq primer: -41m13 fwd. Et from Amersham  
 High quality sequence stop: 363.  
 Location/Qualifiers

FEATURES  
 source

1. 490  
 /organism="Homo sapiens"  
 /db\_xref="GDB:5978049"  
 /db\_xref="taxon:9606"  
 /map="6 p23-p22: 949D05; 2: 2q14.3-2q21.3"  
 /clone="IMAGE:757435"  
 /clone\_11b="Soares\_testis\_NHT"  
 /sex="male"  
 /lab\_host="DH10B"  
 /note="Vector: pT73D-Pac (Pharmacia) with a modified  
 polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA  
 was prepared from mRNA obtained from Clontech  
 Laboratories, Inc., and primed with a Not I - o1190(dT)  
 primer [5].  
 TGTACCAATCGAAGTGGAGCGGCCCAATTTTATTTTATTTT 3'}.  
 Double-stranded cDNA was ligated to Eco RI adaptors  
 (Pharmacia), digested with Not I and cloned into the Not I  
 and Eco RI sites of the modified pT73 vector. Library  
 went through one round of normalization to Cot5, and was  
 constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT  
 ORIGIN

173 a 73 c 101 g 143 t

Query Match 96.2%; Score 352; DB 33; Length 490;  
 Best Local Similarity 99.2%; Pred. No. 7.7e-71;  
 Matches 365; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

2 CTTCTCTCCCTCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTAC 61  
 453 CTTCTCTCTCCCTCTCTGAAATTAATCTTCAACTGCAATTTGCAAGATTAC 394  
 62 ATTTCACGTGATGATATATGTTGC-AAAAAAAAAAAGTCTTTGTTAAATT 118  
 393 ATTTCACGTGATGATATATGTTGC-AAAAAAAAAAAGTCTTTGTTAAATT 334  
 119 ACTGTTTGTGATGATGATGTTTCCCATTTGCAATGATTAACCATCTCT 178  
 333 ACTGTTTGTGATGATGATGTTTCCCATTTGCAATGATTAACCATCTCT 274  
 179 GAACGTGTAAGAAACATCTGAAAGCTAGTCTATCAGCATGTCAGGTAATGGATG 238  
 273 GAACGTGTAAGAAACATCTGAAAGCTAGTCTATCAGCATGTCAGGTAATGGATG 214  
 239 GTTCTCAGAACCATTTTCAACCCAGACAGCTGTTTCTATCCTGTTTAAATTAATTTGG 298  
 213 GTTCTCAGAACCATTTTCAACCCAGACAGCTGTTTCTATCCTGTTTAAATTAATTTGG 154  
 299 GTTCTCTACATGATTAACAACCCCTGCTGCAATGTCATCAATAAAGTGTGACTTGA 358  
 153 GTTCTCTACATGATTAACAACCCCTGCTGCAATGTCATCAATAAAGTGTGACTTGA 94  
 359 GTTACTG 366  
 93 GTTACTG 86

RESULT 6  
 A1685682 423 bp mRNA EST 27-MAY-1999  
 LOCUS ct89h03.x1 NCI\_CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2248757 3',  
 DEFINITION mRNA sequence.  
 ACCESSION A1685682  
 NID 94896976

VERSION A1685682.1 GI:4896976  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 423)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188568.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/HLN at:  
[www.bio.llnl.gov/db/ftp/image/image.html](http://www.bio.llnl.gov/db/ftp/image/image.html)

FEATURES  
source  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1..423  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="19p12-p13.1"  
/clone="IMAGE:2248757"  
/clone\_1lb="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: PT73D-Pac (Pharmacia)  
with a modified polylinker. Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clones  
985608-986759, 1101192-1101959, and 1217928-1220615)."  
Subtraction by Bento Soares and M. Fatima Bonaldo."  
BASE COUNT 150 a 66 c 83 g 124 t  
ORIGIN

Query Match 93.4%; Score 342; DB 50; Length 423;  
Best Local Similarity 100.0%; Pred. No. 1.5e-68;  
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

25 TTTAATCTTTCACTGCAATTTGCAAGATTACATTTCACTGATGATATTTGTG 84  
|||||  
423 TTTAATCTTTCACTGCAATTTGCAAGATTACATTTCACTGATGATATTTGTG 364  
|||||  
85 TTGCAAAAAAAAAAAGTCTTTGTTTAAATTAATCTGTTGTTGTAATCCATCTTCTT 144  
|||||  
363 TTGCAAAAAAAAAAAGTCTTTGTTTAAATTAATCTGTTGTTGTAATCCATCTTCTT 304  
|||||  
145 TTTCCCATTTGAACTGATTAACCATCTGAACTGTAATAAACAATCGAAGAG 204  
|||||  
303 TTTCCCATTTGAACTGATTAACCATCTGAACTGTAATAAACAATCGAAGAG 244  
|||||  
205 CTAGCTATGAGCATCTGACAGTGAATGGATGTTCTGAGAACATTTACCCACAGACA 264  
|||||  
243 CTAGCTATGAGCATCTGACAGTGAATGGATGTTCTGAGAACATTTACCCACAGACA 184  
|||||  
265 GCCTGTTTCTATCTGTTTAAATTAATTAATGTTGGTTCTCTACATGATTAACAACCTCG 324  
|||||  
183 GCCTGTTTCTATCTGTTTAAATTAATTAATGTTGGTTCTCTACATGATTAACAACCTCG 124  
|||||

QY 325 CTCCAATCTGTCACATAAAGTCTGTGACTTGAAGTTTACTC 366  
|||||  
Db 123 CTCCAATCTGTCACATAAAGTCTGTGACTTGAAGTTTACTC 82  
|||||

RESULT 7  
AA229495/c  
LOCUS  
DEFINITION  
AA229495  
NID  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 419)  
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL Unpublished (1997)  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394602.  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chague, M.D.,  
Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution Information can be  
found through the I.M.A.G.E. Consortium/HLN at:  
[www.bio.llnl.gov/db/ftp/image/image.html](http://www.bio.llnl.gov/db/ftp/image/image.html)

Insert Length: 856 Std Error: 0.00  
Seq primer: -41m13 fwd. ET from Amersham  
High quality sequence stop: 368.  
Location/Qualifiers  
1..419  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:1010092"  
/clone\_1lb="NCI-CGAP\_Pr2"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: pAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected preneoplastic cells  
histologically-determined to be prostatic intraepithelial  
neoplasia 2 (PIN2) cells. Double-stranded cDNA was  
ligated to EcoRI adaptors, 5 cycles of PCR applied to the  
cDNA with an adaptor-specific primer, and the resulting  
PCR product subcloned into pAMP10 by the UDG-cloning  
method (Life Technologies). Average insert size is 600  
bp. NOTE: Not directionally cloned. This library was  
constructed by David Kitzman."

BASE COUNT 141 a 68 c 79 g 131 t  
ORIGIN

Query Match 85.5%; Score 313; DB 30; Length 419;  
Best Local Similarity 99.4%; Pred. No. 5.6e-62;  
Matches 325; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 42 GCAATTTGCAAGATTACATTTCACTGTGATGATATTTGTTGC -AAAAAAAAAAAA 99  
|||||  
Db 419 GCAATTTGCAAGATTACATTTCACTGTGATGATATTTGTTGC -AAAAAAAAAAAA 360  
|||||

OY	100	AGTGCCTTGTTAAATTAATTA	CTGTTGGTTGGTGAATGCATCTTGCTTTTCCCATGGAAAC	159
Db	359	AGTGCCTTGTTAAATTAATTA	CTGTTGGTTGGTGAATGCATCTTGCTTTTCCCATGGAAAC	300
OY	160	TAGTCATTAAACCATCTCTGAA	CTGGTAGAAACAATCTGAAGACCTAGTCATACGAT	219
Db	299	TAGTCATTAAACCATCTCTGAA	CTGGTAGAAACAATCTGAAGACCTAGTCATACGAT	240
OY	220	CTGACAGGTGAATTTGGATGG	TTCTCAGAACCATTTTCACCCAGACAGCCTGTTTCATCCT	279
Db	239	CTGACAGGTGAATTTGGATGG	TTCTCAGAACCATTTTCACCCAGACAGCCTGTTTCATCCT	180
OY	280	GTTTAAATTAATTAATTAATTA	AGTTTGGTCTCTACATGCATTAACAACCTGCTCCAACTGTGCACA	339
Db	179	GTTTAAATTAATTAATTAATTA	AGTTTGGTCTCTACATGCATTAACAACCTGCTCCAACTGTGCACA	120
OY	340	TAAAGCTGTGACTGTAAGTTA	CTC 366	
Db	119	TAAAGCTGTGACTGTAAGTTA	CTC 93	
RESULT	8			
LOCUS	AA640928			
DEFINITION	AA640928	411 bp	mRNA	EST 27-OCT-1997
ACCESSION	AA640928			
NID	92566178			
VERSION	AA640928.1	GI:2566178		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 411)			
JOURNAL	NCI-CCGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .			
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
	Tumor Gene Index			
	Unpublished (1997)			
	On Dec 30, 1996 this sequence version replaced g1.1530955.			

## FEATURES

```

Seq primer -28ml3 rev1 ET from Amersham
High quality sequence stop: 387.
Location/Qualifiers
1.411
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1169295"
/clone_1lb="NCI_CGAP_Pr3"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/notes="Vector: PAMPI0; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from
5,000-10,000 microdissected cells
histologically-determined to be fully malignant prostate
cancer cells. Double-stranded cDNA was ligated to EcoRI
adaptors, 5 cycles of PCR applied to the cDNA with an
adaptor-specific primer, and the resulting PCR product
subcloned into pAMP10 by the uDG-cloning method (Life

```

Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krizman."

Query Match	83.28;	Score 304.4;	DB 36;	Length 411;
Best Local Similarity	99.48;	Pred. No. 5e-60;		
Matches 316;	Conservative	0;	Mismatches 1;	Indels 1;
			Gaps	1;

OY	50	CAAGATTCACATTCCTACTGATGATATATGTTGCG-AAAAAAAAAAAGTGCTT	108
Dd	5	CAGAGTTCACATTCCTACTGATGATATATGTTGCGAAAAAAAAAAAGTGCTTT	64
OY	109	GTTTAAATTACTGGTGGTTTGGAATCCATCTTGCTTTTCCCATTTGGAAGTACATTA	168
Dd	65	GTTTAAATTACTGGTGGTTTGGAATCCATCTTGCTTTTCCCATTTGGAAGTACATTA	124
OY	169	ACCCATCTGTAACGTGATGAGAAAAACATCTGAAGAGCTAGTATACAGATGTGCAGGT	228
Dd	125	AACCATCTCTGTAACGTGATGAGAAAAACATCTGAAGAGCTAGTATACAGATGTGCAGGT	184
OY	229	GAATGTGATGGTCTCAGAACCATTTCAACCCAGACAGCGCTGTTCTATCCGTTTAATA	288
Dd	185	GAATGTGATGGTCTCAGAACCATTTCAACCCAGACAGCGCTGTTCTATCCGTTTAATA	244
OY	289	ATTAGTTGGGGTCTCTCATGATCATATAAAACCGTGCATCTGTACATATAAAGTCT	348
Dd	245	ATTAGTTGGGGTCTCTCATGATCATATAAAACCGTGCATCTGTACATATAAAGTCT	304
OY	349	GTGACTTGAAGTTTAGTC 366	
Dd	305	GTGACTTGAAGTTTAGTC 322	

RESULT	9
A1672753/c	
LOCUS	383 bp mRNA EST
DEFINITION	w558d10..x1 Soares_lymus_NHPT.homo sapiens cDNA clone
	18-MAY-1999

ACCESSION	AI672753
NID	94852484
VERSION	AI672753.1
KEYWORDS	GI:4852484
SOURCE	EST.
	human.

ORGANISM	Homo sapiens
REFERENCE	Homaykaya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS	Eutheria; Primates; Carnivora; Homnidae; Homo.
TITLE	1 (bases 1 to 383)
JOURNAL	NCI-CCGAP <a href="http://www.ncbi.nlm.nih.gov/ccgap">http://www.ncbi.nlm.nih.gov/ccgap</a> .
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
	Tumor Gene Index
	Unpublished (1997)
	On May 18, 1998 this sequence version replaced gi:3137861.

## FEATURES

### Source

```

FEATURES
source
    Contact: Robert Strausberg, Ph.D.
    Tel.: (301) 496-1550
    Email: Robert.Strausberg@nih.gov
    This clone is available royalty-free through LNL ; contact the
    IMAGE Consortium (info@image.llnl.gov) for further information.
    Seq primer: -40Up from Gdbco
    High quality sequence stop: 362.
    Location/Qualifiers
        1..383
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /map="952F05; 14; 14q23.1-14q24.1; MMU16C3-C4 region;
            MMU16C3-C4 region"
            /clone="IMAGE:2345299"
            /clone_lib="Soares.thymus_NHFTn"
            /dev_stage="fetal"

```

```

/lab.host="DH10B (phage-resistant)"
/notes="Organ: thymus, pooled; Vector: pRT3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5',
TGTATCCACATCTGAAGTCGAGCGCGCCGACGTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pRT3D vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Fatima Bonaldo. "

```

Query Match	81.7%	Score 299	DB 50	Length 383
Best Local Similarity	100.0%	Pred. No.	8.5e-59	
Matches 299	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Oy	68	CTGTGATCTATATGTTGTCGAAAAAAAAGTGTCTGTTAAATTAATCTGGTTT	127
Db	383	CTGTGATCTATATGTTGTCGAAAAAAAAGTGTCTGTTAAATTAATCTGGTTT	324
Oy	128	GTGAATCCATCTGCTTTTCCCATTTGGAACTAGTCATTAACCCATCTGTGAACGTGA	187
Db	323	GTGAATCCATCTGCTTTTCCCATTTGGAACTAGTCATTAACCCATCTGTGAACGTGA	264
Oy	188	GAAAAACATCTGAAGAGTACTGATATGACATCTGACAGAGGAAATGGAATGGTTCACGA	247
Db	263	GAAAAACATCTGAAGAGTACTGATATGACATCTGACAGAGGAAATGGAATGGTTCACGA	204
Oy	248	ACCAATTTACCCACAGACACCTGTTTCTATCCTGTTTATAAATTAAGTTGGGTCTCTAC	307
Db	203	ACCAATTTACCCACAGACACCTGTTTCTATCCTGTTTATAAATTAAGTTGGGTCTCTAC	144
Oy	308	ATGCATTAACAAACCTGCTCCAAATCTGTCACATATAAAGCTGTGACTGTAAGTTTACTC	366
Db	143	ATGCATTAACAAACCTGCTCCAAATCTGTCACATATAAAGCTGTGACTGTAAGTTTACTC	85

RESULT	10				
LOCUS	AA533772/c				
DEFINITION	AA533772	445 bp	mRNA	EST	21-AUG-1997
ACCESSION	U93605.s1	NCI-CGAP_Fr11	Homo sapiens	CDNA clone	IMAGE:1000040,
NID	AA533772				
VERSION	92277788				
KEYWORDS	AA533772.1	GI:2277788			
SOURCE	EST.				
ORGANISM	human.				
REFERENCE	Homo sapiens				
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
TITLE	Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
	1 (bases 1 to 445)				
	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> .				
	National Cancer Institute, Cancer Genome Anatomy Project (CGAP)				
	Tumor Gene Index				
	Unpublished (1997)				
COMMENT	On May 8, 1995 this sequence version replaced gi:800232.				

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Rodrigo F. Chuaguí, M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Krizman, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bdrip/image/image.html](http://www-bio.llnl.gov/bdrip/image/image.html)

Insert Length: 613      Std Error: 0.00

FEATURES	
source	
	Seq primer: -40m13 fwd. ET from Auerham
	High quality sequence stop: 317.
	Location/Qualifiers
1. .445	
	/orientation="theta_pos" /

BASE COUNT	148 a	69 c	87 g	141 t
ORIGIN				

Query Match	Best Local Similarity	80.4%;	Score 294.2;	DB 34;	Length 445;
Matches 318;	Conservative	98.5%;	Pred. No. 1e-57;	Mismatches 3;	Indels 2;
		0;	Gaps		2;
Qy	45	ATTTCGAAGGATTTACACATTTCACTGATGATATATTGTGTGCACAAAAAAGTGC- 104			
Db	421	AATTGCAAGGATTTACACATTTTCAC-GGGATGATATTTGTGTGCACAAAAAAGTGC 363			
Qy	104	TCTTTGTTAAATTTACTTGGTTGTGATCATCTGTTTTTCCCATTTGGAATCTAGT 163			
Db	362	TCTTTGTTAAATTTACTTGGTTGTGATCAACCATTTCTTTTCCCATTTGGAATCTAGT 303			
Qy	164	CATTACCCATCTCGAAGCTGATAGAAAAACATCGAAGAGTATGATACAGATCTGA 223			
Db	302	CATTACCCATCTCGAAGCTGATAGAAAAACATCTGGAAGAGTATGATACAGATCTGA 243			
Qy	224	CAGTGAATTTGGATGTTCTCAGAACCATTTCCACCAGACGCTGTTTCTATCCCTGTTT 283			
Db	242	CAGTGAATTTGGATGTTCTCAGAACCATTTCCACCAGACGCTGTTTCTATCCCTGTTT 183			
Qy	284	AATAAATTAGTTTGGGTTCTCTACATGCATACAAACCCCTGCTCCAACTGTGCATATAA 343			
Db	182	AATAAATTAGTTTGGGTTCTCTACATGCATACAAACCCCTGCTCCAACTGTGCATATAA 123			
Qy	344	AGTCGTGACTTGAAGTTTAGTC 366			
Db	122	AGTCGTGACTTGAAGTTTAGTC 100			

RESULT	11
AA531606/c	
LOCUS	AA531606
DEFINITION	AA531606 423 bp mRNA EST 20-AUG-1997 J05604.s1 NCI-CGAP_Pri10 Homo sapiens CDNA clone IMAGE:997374, mRNA sequence.
ACCESSION	AA531606
NID	g2274312
VERSION	AA531606.1 GI:2274312
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 423) NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)
JOURNAL	On Sep 12, 1996 this sequence version replaced gi:1395365.
COMMENT	

Contact: Robert Strausberg, Ph.D.  
Tel.: (301) 496-1550  
Email: Robert\_Strausberg@nih.gov  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo F. Chuquibambas

M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbt/p/nci/image/image.html](http://www.bio.llnl.gov/dbt/p/nci/image/image.html)

Insert Length: 897 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 319.  
 Location/Qualifiers

## FEATURES

SOURCE

1. 423  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:997374"  
 /clone\_11b="NCI-CGAP\_Pr10"  
 /sex="male"  
 /tissue\_type="Invasive prostate tumor"  
 /lab\_host="DH10B"  
 /note="Organ: prostate; Vector: pAMP10; mRNA made from  
 Invasive prostate tumor, CDNA made by oligo-dt priming,  
 Non-directionally cloned. Size selected on agarose gel,  
 average insert size 600 bp. Library made by D. Kitzman,  
 NIH."

BASE COUNT 140 a 67 c 82 g 134 t  
 ORIGIN

Query Match 79.8%; Score 292; DB 34; Length 423;  
 Best Local Similarity 97.8%; Pred. No. 3.3e-57;

Matches 317; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

OY 45 ATTGCAAGATTACACATTCAGTGTGATGATGTTGGTGC-AAAAAAAAAAGT 103  
 DB 423 AATTGCAAGATTACACATTCAGTGTGATGATGTTGGTGC-AAAAAAAAAAGT 364  
 OY 104 TCTTTG-TTAAATTAATCTGTTGTGATCCATCTCTTTTCCCATGGAATAG 162  
 DB 363 TCTTTGCTTAAATTAATCTGTTGTGATCCATCTCTTTTCCCATGGAATAG 304  
 OY 163 TCATTACCATCTCTGAACTGTGAGAAAACATCTGAAAGAGTATCATGACATCTG 222  
 DB 303 TCATTACCATCTCTGAACTGTGAGAAAACATCTGAAAGAGTATCATGACATCTG 244  
 OY 223 ACAGGGAATGAGTGTCTCAGAACCATTCACCCAGACGCTGTTTCAATCTCTG 282  
 DB 243 ACAGGGAATGAGTGTCTCAGAACCATTCACCCAGACGCTGTTTCAATCTCTG 184  
 OY 283 TATAAATTAATGTTGGTCTCTACATGATACAAACCCCTGCTCCATCTGACATTA 342  
 DB 183 TATAAATTAATGTTGGTCTCTACATGATACAAACCCCTGCTCCATCTGACATTA 124  
 OY 343 AAGTCTGTGACTGGAAGTTAGTC 366  
 DB 123 AAGTCTGTGACTGGAAGTTAGTC 100

## RESULT 12

AA493522

LOCUS AA493522 403 bp mRNA EST 18-AUG-1997  
 DEFINITION ng75f05.s1 NCI\_CGAP\_Pr6 Homo sapiens CDNA clone IMAGE:940641, mRNA  
 sequence.

ACCESSION

AA493522

VERSION

92223363

KEYWORDS

AA493522.1

GI:2223363

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE

1 (bases 1 to 403)

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT On Jun 18, 1996 this sequence version replaced gi:1366614.

Contact: Robert Strausberg, Ph.D.  
 Tel: (301) 496-1550  
 Email: Robert.Strausberg@nih.gov  
 Tissue Procurement: David B. Boswick, M.D., Rodrigo F. Chuang,  
 M.D., Michael R. Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: David B. Kitzman, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbt/p/nci/image/image.html](http://www.bio.llnl.gov/dbt/p/nci/image/image.html)

## FEATURES

SOURCE

Insert Length: 651 Std Error: 0.00  
 Seq primer: -40ml3 fwd. ET from Amersham  
 High quality sequence stop: 398.  
 Location/Qualifiers

1. 403  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:940641"  
 /clone\_11b="NCI\_CGAP\_Pr6"  
 /sex="male"  
 /tissue\_type="prostate"  
 /lab\_host="DH10B"  
 /note="Vector: pAMP10; mRNA made from prostatic  
 intraepithelial neoplasia (low-grade), CDNA made by  
 oligo-dt priming. Non-directionally cloned.  
 Size-selected on agarose gel, average insert size 600 bp.  
 Reference: Kitzman et al. (1996) Cancer Research  
 56:5380-5383."

BASE COUNT 111 a 85 c 67 g 140 t  
 ORIGIN

Query Match 78.7%; Score 288; DB 34; Length 403;  
 Best Local Similarity 99.3%; Pred. No. 2.7e-56;

Matches 300; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

OY 2 CCTTCTCTCTCCCTCCCTGATTAATCTCTCACTTGAATTCAGAGATTAC 61  
 DB 102 CTTTCTCTCTCCCTCCCTGATTAATCTCTCACTTGAATTCAGAGATTAC 161  
 OY 62 ATTCACGTGTGATGATGTTGTTGC--AAAAAAAAAAGTCTTTGTTAAATTA 119  
 DB 162 ATTCACGTGTGATGATGTTGTTGCAAAAAAAGTCTTTGTTAAATTA 221  
 OY 120 CTGTTTGTGATGATCTTCTTTTCCCATTTGGAACAGTATCATTAACCATCTG 179  
 DB 222 CTGTTTGTGATGATCTTCTTTTCCCATTTGGAACAGTATCATTAACCATCTG 281  
 OY 180 AACTGTAGAAAAACATCTGAAAGAGTGTATCAGCATCTGACAGTGAATGATG 239  
 DB 282 AACTGTAGAAAAACATCTGAAAGAGTGTATCAGCATCTGACAGTGAATGATG 341  
 OY 240 TTCTCAGAACATTTCCACAGACAGCTGTTTCTATCTGTTTAATTAATTAATG 299  
 DB 342 TTCTCAGAACATTTCCACAGACAGCTGTTTCTATCTGTTTAATTAATTAATG 401  
 OY 300 TT 301  
 DB 402 TT 403

## RESULT 13

AA650104

LOCUS AA650104 339 bp mRNA EST 13-NOV-1997  
 DEFINITION ns92f11.s1 NCI\_CGAP\_Pr3 Homo sapiens CDNA clone IMAGE:1191117, mRNA  
 sequence.



ACCESSION AA650104  
NID 92577432  
VERSION AA650104.1 GI:2577432  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 339)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Sep 1, 1995 this sequence version replaced.  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuquib,  
M.D., Michael Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: David B. Kitzman, Ph.D.  
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)  
Insert Length: 1362 Std Error: 0.00  
Seq primer: -40m13 fwd. ET from Amersham  
High quality sequence stop: 291.  
Location/Qualifiers  
1. 339  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="17q21, between D17S1321 and D17S1325"  
/clone IMAGE:1191117  
/clone\_1lb="NCI-CGAP\_Pr3"  
/sex="Male"  
/dev\_stage="45 years old"  
/lab\_host="DH10B"  
/note="Vector: PAMP10; Site\_1: NotI; Site\_2: EcoRI; 1st  
strand cDNA was primed with oligo(dT)17 on 50 ng of  
DNase-treated, total cellular RNA obtained from  
5,000-10,000 microdissected cells  
histologically-determined to be fully malignant prostate  
cancer cells. Double-stranded cDNA was ligated to EcoRI  
adaptors, 5 cycles of PCR applied to the cDNA with an  
adaptor-specific primer, and the resulting PCR product  
subcloned into PAMP10 by the UDG-cloning method (Life  
Technologies). Average insert size is 600 bp. NOTE: Not  
directionally cloned. This library was constructed by  
David Kitzman."

BASE COUNT 94 a 72 c 57 g 116 t  
ORIGIN

Query Match 77.7%; Score 284.2; DB 36; Length 339;  
Best Local Similarity 99.0%; Pred. No. 2e-55;  
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

2 CTCCTTCCTCCCTCCTGGAATTAATCTTCAACTTGCAATTGCAAGATTACAC 61  
|||||  
Db 51 CTCGTCTCCTCCCTCCTGGAATTAATCTTCAACTTGCAATTGCAAGATTACAC 110  
|||||  
62 ATTTCATCTGATATATTTGTTGCAAAAAAGTCTTTGTTAAATTAAT 121  
|||||  
Db 111 ATTTCATCTGATATATTTGTTGCAAAAAAGTCTTTGTTAAATTAAT 170  
|||||  
122 TGGTTTGAATCATCTGCTTTTCCCATGGAATCATTAACCATCTCGAA 181  
|||||  
Db 171 TGGTTTGAATCATCTGCTTTTCCCATGGAATCATTAACCATCTCGAA 230  
|||||  
182 CTGTAGAAAAACATCTGAAAGCTAGTCTATCAGCATCTGACAGTGAATGATGTT 241  
|||||

Db 231 |||||  
CTGTAGAAAAACATCTGAAAGCTAGTCTATCAGCATCTGACAGGGAATGATGTT 290  
Oy 242 CTCAGAACATTTACCCAGACGCTGTTCTATCCTGTTAATAAT 290  
|||||  
Db 291 CTCAGAACATTTACCCAGACGCTGTTCTATCCTGTTAATAAT 339  
|||||

RESULT 14  
AI669511/c 377 bp mRNA EST 14-MAY-1999  
LOCUS wb88f07.x1 NCI-CGAP\_Pr28 Homo sapiens cDNA clone IMAGE:2312773 3',  
DEFINITION mRNA sequence.  
ACCESSION AI669511  
NID 94834285  
VERSION AI669511.1 GI:4834285  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 377)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Dec 20, 1995 this sequence version replaced gi:1130922.  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: [Robert.Strausberg@nih.gov](mailto:Robert.Strausberg@nih.gov)  
Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.  
Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
[www.bio.llnl.gov/dbp/image/image.html](http://www.bio.llnl.gov/dbp/image/image.html)  
Seq primer: -40UP from Gibco.  
Location/Qualifiers  
1. 377  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone IMAGE:2312773  
/clone\_1lb="NCI-CGAP\_Pr28"  
/sex="male"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/note="Organ: prostate; Vector: pTZ193D-Pac (Pharmacia)  
with a modified polylinker; Plasmid DNA from the  
normalized library NCI-CGAP\_Pr22 was prepared, and ss  
circles were made in vitro. Following HAP purification,  
this DNA was used as tracer in a subtractive hybridization  
reaction. The driver was PCR-amplified cDNAs from a pool  
of 5,000 clones made from the same library (clonoids  
985608-986759, 1101192-1101959, and 1217928-1220515).  
Subtraction by Bento Soares and M. Fatima Bonafide."

BASE COUNT 132 a 61 c 74 g 110 t  
ORIGIN

Query Match 76.8%; Score 281; DB 50; Length 377;  
Best Local Similarity 99.7%; Pred. No. 1e-54;  
Matches 292; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

75 GATATGCTGCTGC-AAAAAAAGTCTGTTGTTAAATTAATGTTGTTGAAT 133  
|||||  
Db 377 GATATGCTGTTGCAAAAAAAGTCTGTTGTTAAATTAATGTTGTTGAAT 318  
|||||  
Oy 134 CCAATCTGCTTTTCCCATGGAATCATTAACCATCTCTGAACTGTAAGAAA 193  
|||||

DB	317	CCATCTTGCTTTTCCCATTTGGAAGTACGATTAACCCATCTCTGTGAACGTGTAGAAAAA	258
QY	194	CATCTGAAAGAGCTAGCTCTATCAGCATCTGACAGAGTGATGGATGGTCTCAGAACCAT	253
Db	257	CATCTGAAAGAGCTAGCTCTATCAGCATCTGACAGAGTGATGGATGGTCTCAGAACCAT	198
QY	254	TACACCAGACAGCGCTTTTCTATCCTGTGTAAATAAATAGTTGGGCTTCCTCATGCAT	313
Db	197	TCACCCAGACAGCGCTTTTCTATCCTGTGTAAATAAATAGTTGGGCTTCCTCATGCAT	198
QY	314	AACAAACCTGCTCCATCTGTGCATTAATAAGCTGTGACTTGAAGTTAGTC	366
Db	137	AACAAACCTGCTCCATCTGTGCATTAATAAGCTGTGACTTGAAGTTAGTC	85
RESULT 15			
LOCUS	W47380	374 bp	mRNA
DEFINITION	zc39h11.s1 Soares, senescent, fibroblasts, NBHSF Homo sapiens cDNA, clone IMAGE:324741 3, similar to contains Alu repetitive element, mRNA sequence.		
ACCESSION	W47380		
VERSION	W47380.1	GI:1332019	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 374) Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucada, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.		
TITLE	The Mashu-Merck EST Project		
JOURNAL	Unpublished (1995)		
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@wustl.edu This clone is available royalty-free through LMD; contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: mob.BEAG-ET High quality sequence stop: 360. Location/Qualifiers 1..374 /organism="Homo sapiens" /db_xref="GDB:1256253" /db_xref="taxon:9606" /clone="IMAGE:324741" /clone_1ib="Soares, senescent, fibroblasts, NBHSF" /tissue_type="senescent fibroblast" /lab_host="DH10B (ampicillin resistant)" /note="Vector: pTZ19 (Pharmacia) with a modified polylinker V-type; phagemid; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer 15' TGTTGCAATCTGAAGTGGAGCGCCGCGCATTTTTTTTTTTTTTTT 3', double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pTZ19 vector (Pharmacia). Library was through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M. Felina Bonaldo."		
BASE COUNT	131 a	59 c	74 g 110 t
ORIGIN			

	Best Local Similarity	99.7%	Pred. No. 1.8e-54;	Matches	291;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
QY	76	TATATTGTTGTC	AAAAAAAAAAAGTGTCTTTGTTAAATTA	CTGTTGTTGTAATC	134								
Db	374	TATATTGTTGCCAAAA	AAAAAAAAAAGTCTTTGTTAAATTA	CTGTTGTTGTAATC	315								
QY	135	CATCTGCTTTTCCC	CACTGCACTAGCATTA	CCCATCTCAACGCTGAGAAAAAC	194								
Db	314	CATCTGCTTTTCCC	CATGCACTAGCATTA	CCCATCTCTCAACGCTGAGAAAAAC	255								
QY	195	ATCTAAGAGCTAGCT	ATACGATCTACAGAGTAA	TGATGCTTCACAGACATT	254								
Db	254	ATCTAAGAGCTAGCT	ATACGATCTACAGAGTAA	TGATGCTTCACAGACATT	195								
QY	255	CACCCAGACAGCT	TTTCTATCTGTTTAA	TAAATAGTTGGTCTCTACATGACATA	314								
Db	194	CACCCAGACAGCT	TTTCTATCTGTTTAA	TAAATAGTTGGTCTCTCTACATGACATA	135								
QY	315	ACAACCCCTGCTCC	CAATGCTGCATTA	AAAGCTGTGACTGAATTTAGTC	366								
Db	134	ACAACCCCTGCTCC	CAATGCTGCATTA	AAAGCTGTGACTGAATTTAGTC	83								

Search completed: September 25, 1999, 12:05:07  
Job time: 8111 sec

## Query-Match

76.58; Score 280; DB 26; Length 374;

---

**This Page Blank (uspto)**



```
56 LeuGIuAlaAspGInGUProGlySerGlnMetValGluAlaSerLeu 72
198 CGTACGGCACCAGAGTAAACAGACACCTTGTGCTGCTGACAGCTCATGC 247
72 rValArgHisProGluIuTrpAsnArgProLeuLeuAlaAsnAspLeuMetL 89
248 TCATCAAGTTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGCATC 297
89 euriLeuIserAspGluSerValSerGluSerAspThrIleArgSerIle 105
298 AGCATTTGCTTGCAGTGCCTTACCGCGGGGAGACTTGTGCTTCTTCTGG 347
106 SerIleAlaSerGInGlyProThrAlaGlyAsnSerGlyLeuValSerGI 122
348 CTGGGGTGTGCTGGCAACGGGTGAGCTCAGCGGGTGTGTGCTGCCCTCT 397
122 YTPGlyLeuLeu..... 126
398 CAAGGAGGTCTCTGTCCAGTCCGCGGGGCTGACCCAGAGCTGTGCTTC 447
127 .....AlaAsn 128
448 CAGCGAGAAATGCTTACCGTGTGCTGAGTGTGAGTGTGCTGTCT 497
129 ..GlyArgMetProThrValLeuGInGlyValAlaAsnValSerValSer 144
498 GAGGAGTCTGCAGTAACTTATGACCCGCTGATACACCCAGCATAGTT 547
145 GInGIuValGlySerIuSerIuSerIuSerIuSerIuSerIuSerIuSer 161
548 CTGCGCGCGCGGAGGCAAGACAGAAAGAGTCTGCAACGGTACTCTG 597
161 ecyAlaIleGlyGlyGlnAspGlnIuAspSerCysAsnGlyAspSerG 178
598 GGGGGCCCTGATGTGCAACGGGTACTTGCAGGCGCTTGTCTTCTTGG 647
178 IyGlyProLeuIleCysAsnGlyIuTrpLeuGInGlyLeuValSerPhe 194
648 AAACGCCCGTGTGGCAAGTGGGCTGCGAGTGTACACCAACACTCTG 697
195 LysAlaProCysGlyGlnValGlyValProGlyValIuTrpAsnLeuCy 211
698 CAATTCAGTACTGATGATAGAGAAACCGT.CCAGGCCAGTTAATCTGG 746
211 slyPheThrGluTrpIleGlyIuTrpIleValProGlyGlnLeuThrLeu 228
747 GGACTGGGAACCA 760
228 IyThrGlyAsnPro 232
seq_name: A_Geneseq_36:W60592
seq_documentation_block:
ID W60592 standard; Protein: 248 AA.
AC W60592;
DE 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) protein.
KW Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
KW benign prostate hyperplasia; diagnosis; drug screening; PSK.
OS Homo sapiens.
FH Key Location/Qualifiers
FT MISC_difference 113
FT /label= unknown
FT /note= "encoded by NTC"
FT MISC_difference 128
FT /label= unknown
FT /note= "encoded by AGN"
FT MISC_difference 132
FT /label= unknown
FT /note= "encoded by GNT"
PD WO9820117-A1.
PD 14-MAY-1998.
PF 31-OCT-1997; U20051.
```

```
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goli, SK;
DR WPI: 98-286933/23.
DR N-PSDB: V37495.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 1: Fig 1A-C; 68bp; English.
CC This represents a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 248 AA:
```

```
alignment_scores:
Quality: 1081.50 Length: 243
Ratio: 5.077 Gaps: 2
Percent Similarity: 87.654 Percent Identity: 86.008
```

```
alignment_block:
US-09-030-606-173 x W60592 ..
```

```
Align seg 1/1 to: W60592 from: 1 to: 248
```

```
10 CACTCGAGCCCTGGCAGGCGGACATGATGAAAGCATTTGTTCTG 59
34 HisSerIleProThrPrgInAlaAlaLeuValMetGluAsnIuLeuPheCy 50
60 CTCGGGCGCTGTCGATCCGAGTGGGTGTGTCAGCCGACACTGTT 109
50 sSerGlyValLeuValHisProGlnTrpValIuSerAlaAlaHisCysP 67
110 TCAGAACTCTTACACCATCGGCGCTGGCTGCACAGTCTTGGAGCCGAC 159
67 heGlnAsnSerIyThrIleGlyLeuGlyLeuHisSerLeuGluAlaAsp 83
160 CAGAGCCAGGAGCCAGATGTGTGGAGCCAGCTCTCGTACGGCACC 209
84 GlnGluProGlySerGlnMetValGluAlaSerLeuSerValArgHisP 100
210 AGAGTACACAGACCCCTTGTGCTGCTAAGCAGCTCATGCTCATCAAGTTGG 259
100 oGluTrpAsnArgProLeuLeuAlaAsnAspLeuMet***IleLeuLeu 117
260 ACGAATCCGTGTCCGAGTCTGACACCATCCGAGCATCAGCATGCTTGC 309
117 spGluSerValSerGluSerAspAsnIleArg***IleSerIle***Ser 133
310 CAGTGGCTTACCGCGGGAACCTTGTGCTGTTTGTGCTGGGTGCTCT 359
134 GlnCysProThrAlaGlyAsnPheCysLeuValSerGlyIyPrlGlyLeu 150
360 GCGAAGCGGTGACTCAGCGGGTGTGTGTGCTGCCCTTCAAGAGAGTCT 409
150 u..... 150
410 CTGCCAGTCCGCGGGGCTGACCCAGAGCTGTGCTCCAGCGAGATGC 459
151 .....AlaAsn.. GlyArgMetP 156
460 CTACCGTGTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 509
156 roThrValLeuGInGlyValAlaSerValSerValSerGluGluValCys 172
510 AGTAGCTTATGACCCGCTGTACCAACCCAGCATGTTTGGCCGCGCG 559
```

```

|||||
173 SerLysLeuTYrAspProLeuTYrHisProSerMetPheCysAlaGlyL 189
560 AGGCGACAGACCAAGAGACTCTGCAACGGTGACTCTGGGGGCCCTTGA 609
189 YGlyGlnAspGlnLysAspSerCysAsnGlyAspSerGlyLProLeuI 206
610 TCTGCAACGGGTACTTGCAGGGCCCTGTGTCTTTCGAAAGCCCGTGT 659
206 IecYsAsnGlyTYrLeuGlnGlyLeuValSerPheGlyLysAlaProCys 222
660 GCGCAAGTGGCGGTGCGAGGTGTCTACACCAACTCTCAATTCACATGA 709
223 GlyGlnValGlyValProGlyValTYrThrAsnLeuGlyLysPheThrG 239
710 GTGATAGAGAAAACGTCACAGCCACT 737
239 uTrpIleGlyLysThrValGlnAlaSer 248

```

seq\_name: A\_Geneseq\_36:W69388

seq\_documentation\_block:

```

ID W69388 standard; Protein: 205 AA.
AC W69388;
DE 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE13 protein.
KM Prostate tumour specific gene; human; prostate cancer; detection:
therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT MISC_difference 127 /note= "unspecified amino acid"
FT MISC_difference 204 /note= "unspecified amino acid"
FT W09837418-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (COR1-) CORIXA CORP.
PA Dillon DC, Xu J;
PI WPI: 98-480805/41.
DR N-PSDB: V58647.
PT Novel human prostate specific tumour protein and fragments - useful
PT for detecting and treating prostate cancers
PS Example 1; Page 115-116; 141pp; English.
CC This sequence is encoded by a human prostate tumour specific gene, and
CC can be used in the method of the invention. The method is for detecting
CC prostate cancer comprises contacting a biological sample with an agent
CC able to bind an immunogenic portion of a prostate protein (such as
CC this protein sequence). An antibody which binds to an immunogenic
CC portion of the prostate protein, and the method can be used to detect,
CC monitor progression of, or treat prostate cancers. The antibody may
CC also be conjugated to a therapeutic agent for use in therapy of prostate
CC cancers.
SQ Sequence 205 AA;

```

Alignment\_scores:

```

Quality: 1021.50 Length: 233
Ratio: 5.032 Gaps: 2
Percent Similarity: 87.124 Percent Identity: 85.408

```

Alignment\_block:

US-09-030-606-173 x W69388 ..

Align seg 1/1 to: W69388 from: 1 to: 205

```

40 ATGCAAAAGCAATTGTTCTGCTGGGCGCTCTGTCATCCGACAGTGGT 89
|||||
1 MetGlnAsnGlnLeuPheCysSerGlyValLeuValHisProGlnTrpVal 17

```

```

90 GCTGTACAGCCGACACTGTGTTCCAGAACTCCTACACCATCGGCTGGCC 139
|||||
17 IleuSerAlaIleHisCysPheGlnAsnSerTYrThrIleGlyLeuGlyL 34
140 TGCAAGCTTTGAGCGCCGACCAAGAGCCAGGAGCCAGATGTGGAGGCC 189
34 euHisSerLeuGlnAlaAspGlnGlyProGlySerGlnMetValGlnAla 50
190 AGCTTTCCTGACGGCAGCCAGAGTACACAGACCTTGTCTGCTTAACGA 239
51 SerLeuSerValArgHisProGlyTYrAsnArgLeuLeuAlaAsnAs 67
240 CCTATGCTCATCAAGTTGGAGCAATCGGTGCTCCGAGTCTACACATCC 289
67 PheMetLeuIleLysLeuAspGlnSerValSerGlnSerAspThrIle 84
290 GGACATCAGCATTCGTTGCGAGTGCCTACCGGCGGGAACCTTGCCTC 339
84 rGserIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysLeu 100
340 GTTCTGTGCTGGGCTGTCTGTGCGAAGCGGTAGCTCACGGGTGTGTCT 389
101 ValSerGlyTrpGlyLeuLeu..... 107
390 GCCCTCTTCAAGAGAGTCTCTGCCAGATCGCGGGGCTGACCCAGAGCT 439
107 ..... 107
440 CTGCGTCCGACGAGAAATGCTTACCGTGTCTGACGTGTAACGTGCGG 489
108 ..AlaAsn..GlyArgMetProThrValIleuHisCysValAlaAsnValSerV 123
490 TGGTGTCTGAGAGAGTGTGCGAGTAAGCTTTGACCCGCTTACACCCC 539
123 aValSerGln**ValCysSerLysLeuTYrAspProLeuTYrHisPro 139
540 AGCATGTCTGTCGCGCGGCGGAGGCGCAAGACCAAGAGACTCTGCAACG 589
140 SerMetPheCysAlaGlyGlyGlnAspGlnLysAspSerCysAsnG 156
590 TGACTCTGGGGGCGCCCTGATCTGCAACGGGTACTTGCAGGCGCTTGTGT 639
156 YAspSerGlyLProLeuIleCysAsnGlyTYrLeuGlnGlyLeuValS 173
640 CTTTGGCAAAAGCCCGTGTGCGCAAGTGGCGGTGCCAGGTGTATACAC 689
173 erPheGlyLysAlaProCysGlyLeuGlyValProGlyValTYrThr 189
690 AACCTGTCAAAATTCACGTGATGATAGAGAAACCGTCAGGCGCAGT 737
190 AsnLeuCysLysPheThrGlnTrpIleGlyLysThrValGln***Ser 205

```

seq\_name: A\_Geneseq\_36:W71872

seq\_documentation\_block:

```

ID W71872 standard; Protein: 205 AA.
AC W71872;
DE 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DE13.
KM Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT MISC_difference 127 /note= "undefined residue"
FT MISC_difference 204 /note= "undefined residue"
FT W09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998; U03492.
PR 09-FEB-1998; US-020956.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (COR1-) CORIXA CORP.

```

PI Dillon DC, Xu J;  
 DR WPI: 98-609886/51.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 PS used in a vaccine for the treatment of prostate cancer  
 CC Example 3: Page 107-108; 130pp; English.  
 CC The present sequence is an immunogenic portion of a prostate tumour  
 CC protein. The immunogen, or the DNA encoding it, can be used as a  
 CC vaccine for the treatment of prostate cancer. The immunogen was  
 CC isolated from a prostate tumour cDNA library obtained by subtracting  
 CC a prostate tumour cDNA expression library with a normal tissue cDNA  
 CC library.  
 SO Sequence 205 AA:

alignment\_scores:  
 Quality: 1021.50 Length: 233  
 Ratio: 5.032 Gaps: 2  
 Percent Similarity: 87.124 Percent Identity: 85.408

alignment\_block:  
 US-09-030-606-173 x W71872 ..

Align seg 1/1 to: W71872 from: 1 to: 205

```

40 ATGGAACGAAATTTCTCTCGGCGCTCTGTCATCCGAGTGGT 89
|||||
1 MetGlnAsnGlnLeuPheCysSerGlyValLeuValHisProGlnTrpVal 17
90 GCTGTACGCCGACACTGTTTCCGAACCTCTACACATGGGCGTGGCC 139
|||||
17 LeuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuLysL 34
140 TGCACACTTGTGAGGCGCAGCAAGAGCGAGGAGGAGAGTGGTGGAGGCC 189
|||||
34 euHisSerLeuGlnAlaAspGlnGluProGlySerGlnMetValGlnAla 50
190 AGCCTCTCCGTAGCGACCCAGAGTACAGACAGCCCTGTGCTGCTAAAGA 239
|||||
51 SerLeuSerValArgHisProGlnTyrAsnArgLeuLeuLeuAlaAsnAs 67
240 CTTATCTCTATCAAGTGTGAGCAAGTCCGTTGCTGAGTCTGACACATCC 289
|||||
67 PheMetLeuIleLeuLeuAspGlnSerValSerGlnSerAspThrIleA 84
290 GAGACATCAGCATGCTTCCAGTCCCTACCGCGGGGAACTTGTGCTC 339
|||||
84 rGserIleSerIleAlaSerGlnCysProThrAlaGlnAsnSerCysLeu 100
340 GTTTCGTGGGTGGGTCTGTGGCGAACGGTGAAGCTACGGGTGTGTCT 389
|||||
101 ValSerGlyTrpGlyLeuLeu..... 107
390 GCCCTCTCAAGAGGTCCTCTGCCCACTCGCGGGGCTGACCAAGACT 439
107 ..... 107
440 CTGGTCCGAGGAGATGCTTACGCTGAGTGCAGTGCAGTGCAGTGCAG 489
|||||
108 ..AlaAsn..GlyArgMetProThrValLeuHisCysValAsnValSerV 123
490 TGGTGTGAGGAGTGTGAGTAAAGCTTATGACCCGCTGTACACCCC 539
|||||
123 aValSerGlu**ValCysSerLysLeuTyrAspProLeuTyrHisPro 139
540 AGCATGTTCTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 589
|||||
140 SerMetPheCysAlaGlyGlyGlnAspGlnLysAspSerCysAsnG 156
590 TGACTGTGGGGGGCGCTGATCTGCAAGGGGTATTCAGAGGCGCTTGTGT 639
|||||
156 YAspSerGlyGlyProLeuIleCysAsnGlyTyrLeuGlnGlyLeuValS 173
640 CTTTCGGAAGAGCCCGTGTGGCCAACTTGCGCGTGCAGGTGTCTACAC 689

```

```

|||||
173 ePheGlyLysAlaProCysGlyGlnLeuGlyValProGlyValTyrThr 189
690 AACCTGTCAAATTCATGATGATAGAGAGAAACCGTCCAGGCCACT 737
|||||
190 AsnLeuCysLysPheThrGlnTrpIleGlnTyrThrValGln**Ser 205
seq_name: A_Geneseq_36:W69387

```

seq\_documentation\_block:  
 ID W69387 standard; Protein: 159 AA.

DE 08-DEC-1998 (first entry)  
 DE Prostate tumour specific gene clone DEL protein.  
 KW Prostate tumour specific gene; human; prostate cancer; detection;  
 KW therapy.  
 OS Homo sapiens.

FT Key Location/Qualifiers

FT MISC\_difference 103

FT MISC\_difference 105 /note= "unspecified amino acid"

FT MISC\_difference /note= "unspecified amino acid"

PN W09837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998; U03690.

PR 09-FEB-1998; US-904809.

PR 25-FEB-1997; US-806596.

PR 01-AUG-1997; US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J;

DR WPI: 98-480805/41.

DR N-PSDB: V58644.

PT Novel human prostate specific tumour protein and fragments - useful  
 for detecting and treating prostate cancers

PS Example 1: Page 112-113; 141pp; English.

CC This sequence is encoded by a human prostate tumour specific gene, and  
 can be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprises contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC this protein sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.

SO Sequence 159 AA:

alignment\_scores:  
 Quality: 786.50 Length: 187  
 Ratio: 5.010 Gaps: 2  
 Percent Similarity: 83.957 Percent Identity: 83.422

alignment\_block:  
 US-09-030-606-173 x W69387 ..

Align seg 1/1 to: W69387 from: 1 to: 159

```

178 ATGGTGGAGGCGACCTCTCCGTAGCGGACCCAGAGTATACAGACCTT 227
|||||
1 MetValGlnAlaSerLeuSerValArgHisProGlnTyrAsnArgProLe 17
228 GTCGCGTAAGACCTCATGCTCATCAAGTGTGAGCAATCGGTGCGAGT 277
|||||
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspGlnSerValSerGlnS 34
278 CTGACACCATCCGAGCATCAGCATGCTTCCGAGTGCCTTACCGCGGG 327
|||||
34 eAspThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
328 AACTCTTGGCTCTTCTGTGCTGGGTCTGCTGCGCAAGCGTGAAGTCA 377
|||||
51 AsnSerCysLeuValSerGlyTrpGlyLeu..... 61

```



```

378 GGGTGTGTCTGCCCTTCAAGAGGTCTCTGCCAGTGGCGGGGC 427
61 ..... 61
428 TGACCCAGAGCTCTGCTCCAGGAGAGTAATGCTTACCGTCTGACGTGG 477
62 .....Alaasn..GlyArgMetProThrValLeuGlnCysV 73
478 TGAACGTGTGGTGTCTGAGAGGTCTGAGTAAGCTTATGACCCG 527
73 AlaSnValSerValValSerGlnGlnValCysSerLysLeuThrAspPro 89
528 CTGTACCAACCCAGCATGTTCTGCGCCGCGGAGGAGGCAAGACCAAGA 577
90 LeuThrHisProSerMetPheCysAlaGlyGlyGln***Gln**As 106
578 CTCCTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTTC 627
106 pSerCysAsnGlyAspSerGlyGlyProLeuIleCysAsnGlyTyrLeuG 123
628 AGGGCTGTGTCTTGGAAAGCCCGTGTGGCAAGTTGGCGTCCA 677
123 lnglLeuValSerPheGlyLysAlaProCysGlyGlnValGlyAlPro 139
678 GGTGTCTACACCACTCTGCAATTCATCTGAGTGAATAGAAAACCGT 727
140 GlyValTyrThrAsnLeuCysLysPheThrGluTyrPileGluLysThrVa 156
728 CCAGGCCAGT 737
156 lGlnAlaSer 159

seq_name: A_Geneseq_36:W71871
seq_documentation_block:
ID W71871 standard; Protein: 159 AA.
AC W71871;
DT 06-JAN-1999 (first entry)
DE Protein encoded by prostate tumour clone P703 splice variant DEL.
KW Prostate; cancer; tumour; vaccine; immunogen; clone.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 103 /note= "undefined residue"
FT Misc_difference 105 /note= "undefined residue"
FT W09837093-A2.
PN 27-AUG-1998.
PD 25-FEB-1998; US-020956.
PR 09-FEB-1998; US-806099.
PR 25-FEB-1997; US-806099.
PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J.
DR WPI: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PS used in a vaccine for the treatment of prostate cancer
PS Example 3; Page 105; 130pp; English.
CC The present sequence is an immunogenic portion of a prostate tumour
CC protein. The immunogen, or the DNA encoding it, can be used as a
CC vaccine for the treatment of prostate cancer. The immunogen was
CC isolated from a prostate tumour cDNA library obtained by subtracting
CC a prostate tumour cDNA expression library with a normal tissue cDNA
CC library.
SQ Sequence 159 AA;

```

```

US-09-030-606-173 x W71871 ..
Align seg 1/1 to: W71871 from: 1 to: 159
178 ATGTGTGAGGCCAGCTCTCCGTACGGCAGCCAGAGTACAAACAGACCTT 227
1 MetValGlnAlaSerLeuSerValArgHisProGluTyrAsnArgProLe 17
228 GCTGCTTACCAACCTCATGCTCATCAAGTTGAGCAATCCGTCCGAGT 277
17 uLeuAlaAsnAspLeuMetLeuIleLysLeuAspLeuSerValSerGln 34
278 CTGACACCATCCGAGCATCATGACATTCCTTGCAGTGCCTACCGCGGG 327
34 eAspPheThrIleArgSerIleSerIleAlaSerGlnCysProThrAlaGly 50
328 AACTCTTCCCTGCTGTTCTGGCTGGGGGTCTGCTGGCAAGGTGAGCTCAC 377
51 AsnSerCysLeuValSerGlyTyrPglyLeuLeu..... 61
378 GGGTGTGTCTGCCCTTCAAGAGGTCTCTGCCAGTGGCGGGGC 427
61 ..... 61
428 TGACCCAGAGCTCTGCTCCAGGAGAGTAATGCTTACCGTCTGACGTGG 477
62 .....Alaasn..GlyArgMetProThrValLeuGlnCysV 73
478 TGAACGTGTGGTGTCTGAGAGGTCTGAGTAAGCTTATGACCCG 527
73 AlaSnValSerValValSerGlnGlnValCysSerLysLeuThrAspPro 89
528 CTGTACCAACCCAGCATGTTCTGCGCCGCGGAGGAGGCAAGACCAAGA 577
90 LeuThrHisProSerMetPheCysAlaGlyGlyGln***Gln**As 106
578 CTCCTGCAACGGTGAATCTGGGGGGCCCTGATCTGCAACGGGTACTTC 627
106 pSerCysAsnGlyAspSerGlyGlyProLeuIleCysAsnGlyTyrLeuG 123
628 AGGGCTGTGTCTTGGAAAGCCCGTGTGGCAAGTTGGCGTCCA 677
123 lnglLeuValSerPheGlyLysAlaProCysGlyGlnValGlyAlPro 139
678 GGTGTCTACACCACTCTGCAATTCATCTGAGTGAATAGAAAACCGT 727
140 GlyValTyrThrAsnLeuCysLysPheThrGluTyrPileGluLysThrVa 156
728 CCAGGCCAGT 737
156 lGlnAlaSer 159

seq_name: A_Geneseq_36:W69389
seq_documentation_block:
ID W69389 standard; Protein: 164 AA.
AC W69389;
DT 08-DEC-1998 (first entry)
DE Prostate tumour specific gene clone DE14 protein.
KW Prostate tumour specific gene; human; prostate cancer; detection;
KW therapy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 118 /note= "unspecified amino acid"
FT W09837418-A2.
PN 27-AUG-1998.
PD 25-FEB-1998; US-03690.
PR 09-FEB-1998; US-904809.
PR 25-FEB-1997; US-806596.
PR 01-AUG-1997; US-904809.
PA (CORI-) CORIXA CORP.
PI Dillion DC, Xu J.

```

DR MPI: 98-480805/41.  
 DR N-PSDB: V58648.  
 PT Novel human prostate specific tumour protein and fragments - useful  
 PT for detecting and treating prostate cancers  
 PS Example 1: Page 117-118, 141pp; English.  
 CC This sequence is encoded by a human prostate tumour specific gene, and  
 CC can be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprising contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC this protein sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 SQ Sequence 164 AA:

alignment\_scores:      Quality: 579.00      Length: 165  
                          Ratio: 4.420      Gaps: 3  
 Percent Similarity: 79.394      Percent Identity: 73.333

alignment\_block:  
 US-09-030-606-173 x W69389 ..

Align seg 1/1 to: W69389 from: 1 to: 164

```

40 ATGGAAGCAATGTCTCTCGGGCTCTGGAGCATCCGAGTGGGT 89
|||||
1 Metclunsnluleuphncysserglyvalleuvalhispoglntrpva 17
90 GCTGTCAAGCCGACACTGTTCCAGAACTCTACACATCGGGGTGGCC 139
|||||
17 lleuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
140 TGCACAGTCTTGAGCGCCAGCAAGAGCCAGGAGCCAGATGTGGAGCC 189
|||||
34 euHISserleuGlnAlaAspGlnIuprogIyserGlnMetValGlnAla 50
190 AGCTTCGCGTACGGACCCAGATACAGACCCCTGCTCGCTACAGA 239
|||||
51 SerleuSerValArgHisPProGlnIutyrAsnArgProleuLeuAlaAsnAs 67
240 CCTCATGCTCATCAAGTTGAGAGATCCGCTCGAGTCTGACACATCC 289
|||||
67 pleuMetleuIleIysleuAspGlnSerValSerGlnSerAspThrIleA 84
290 GGAGCATCAGATTGCTTGGCAGTGCCTACCGGGGGGAACTTTGCCTC 339
|||||
84 rgserrIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysleu 100
340 GTTCTGCGTGGGGTCTGCTGGCGAAGCTAGCTACGGGTGTG.... 384
|||||
101 ValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAlaIleGlnse 117
385 .....TGTCTGCCCTCTTCAAGAGAGTCTCTG 412
|||||
117 r**ThrValGlyIleTrpGlnCys.....GlnIysleuSerG 130
413 CCCAGTCCGGGGGCTACCCAGAGCTTGGTCCCGAGGAGAAATGCTTA 462
|||||
130 InProTrpGlnGlyCysThrIleSerAlaThrSerSerAlaArg...Thr 145
463 CCGGTGTCAGATGGGTGAGAGTGCAGTGTCTGAGAGAGTCTC 507
|||||
146 SerCysCysIleLeuThrGlyCysSerleuLeuThrAlaSer 160
seq_name: A.Geneseq_36:W71873

```

seq\_documentation\_block:  
 ID W71873 standard: Protein: 164 AA.  
 AC W71873;  
 DT " 06-JAN-1999 (first entry)

DE Protein encoded by prostate tumour clone P703 splice variant DE14.  
 KW Prostate; cancer; tumour; vaccine; immunogen; clone.  
 OS Homo sapiens.  
 FH Key      Location/Qualifiers  
 FT Misc.difference 118  
 PI      /note= "undefined residue"  
 PT WO9837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998: U03492.  
 PR 09-FEB-1998: US-020956.  
 PR 25-FEB-1997: US-806099.  
 PR 01-AUG-1997: US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR MPI: 98-609886/51.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 PT used in a vaccine for the treatment of prostate cancer  
 PS Example 3: Page 109; 130pp; English.  
 CC The present sequence is an immunogenic portion of a prostate tumour  
 CC protein. The immunogen, or the DNA encoding it, can be used as a  
 CC vaccine for the treatment of prostate cancer. The immunogen was  
 CC isolated from a prostate tumour cDNA library obtained by subtracting  
 CC a prostate tumour cDNA expression library with a normal tissue cDNA  
 CC library.  
 SQ Sequence 164 AA:

alignment\_scores:      Quality: 579.00      Length: 165  
                          Ratio: 4.420      Gaps: 3  
 Percent Similarity: 79.394      Percent Identity: 73.333

alignment\_block:  
 US-09-030-606-173 x W71873 ..

Align seg 1/1 to: W71873 from: 1 to: 164

```

40 ATGGAAGCAATGTCTCTCGGGCTCTGGAGCATCCGAGTGGGT 89
|||||
1 Metclunsnluleuphncysserglyvalleuvalhispoglntrpva 17
90 GCTGTCAAGCCGACACTGTTCCAGAACTCTACACATCGGGGTGGCC 139
|||||
17 lleuSerAlaAlaHisCysPheGlnAsnSerTyrThrIleGlyLeuGlyL 34
140 TGCACAGTCTTGAGCGCCAGCAAGAGCCAGGAGCCAGATGTGGAGCC 189
|||||
34 euHISserleuGlnAlaAspGlnIuprogIyserGlnMetValGlnAla 50
190 AGCTTCGCGTACGGACCCAGAGTACAGACCCCTGCTCGCTACAGA 239
|||||
51 SerleuSerValArgHisPProGlnIutyrAsnArgProleuLeuAlaAsnAs 67
240 CCTCATGCTCATCAAGTTGAGAGATCCGCTCGAGTCTGACACATCC 289
|||||
67 pleuMetleuIleIysleuAspGlnSerValSerGlnSerAspThrIleA 84
290 GGAGCATCAGATTGCTTGGCAGTGCCTACCGGGGGGAACTTTGCCTC 339
|||||
84 rgserrIleSerIleAlaSerGlnCysProThrAlaGlyAsnSerCysleu 100
340 GTTCTGCGTGGGGTCTGCTGGCGAAGCTAGCTACGGGTGTG.... 384
|||||
101 ValSerGlyTrpGlyLeuLeuAlaAsnAspAlaValIleAlaIleGlnse 117
385 .....TGTCTGCCCTCTTCAAGAGAGTCTCTG 412
|||||
117 r**ThrValGlyIleTrpGlnCys.....GlnIysleuSerG 130
413 CCCAGTCCGGGGGCTACCCAGAGCTTGGTCCCGAGGAGAAATGCTTA 462
|||||
130 InProTrpGlnGlyCysThrIleSerAlaThrSerSerAlaArg...Thr 145

```

463 CCGTCTCAGTCCGTGAACGTGCTGCTCTGAGAGGCTCT 507  
 146 SerCysCysIleLeuThrGlyCysSerLeuLeuThrAlaSer 160  
 seq\_name: A\_Geneseq\_36:W94493

seq\_documentation\_block:  
 ID W94493 standard; Protein: 268 AA.  
 AC W94493;  
 DT 23-APR-1999 (first entry)  
 DE Human kallikrein.  
 KW Human; kallikrein; keratinocyte; HKALL; skin disorder; cancer; eczema;  
 KM psoriasis; scleroderma; adenocarcinoma; leukaemia; melanoma.  
 OS Homo sapiens.  
 PN MO9842849-A1.  
 PD 01-OCT-1998.  
 PF 25-MAR-1998; U05939.  
 PR 26-MAR-1997; US-824874.  
 PA (INCY-) INCYTE PHARM INC.  
 PI Hillman JL, Lai P;  
 DR WPI: 99-070073/06.  
 DR N-PSDB: X16295.  
 PT Human kallikrein polypeptide, HKALL - useful e.g. to treat skin  
 PT disorders e.g. eczema, psoriasis and screen for antagonists useful  
 PT to treat skin disorders and cancers  
 PS Claim 1: Page 41-42: 61pp: English.  
 CC The present sequence represents human kallikrein, designated HKALL.  
 CC HKALL (or fragments) can be administered therapeutically to increase  
 CC proteolysis and subsequent skin scaling to treat/prevent skin disorders,  
 CC e.g. eczema, psoriasis and scleroderma. HKALL has chemical and  
 CC structural homology with human stratum corneum chymotryptic enzyme  
 CC (thought to be involved in the proteolysis of intercellular cohesive  
 CC structures necessary for desquamation, the process by which outer layers  
 CC of skin are eliminated), and its expression in cDNA libraries was  
 CC associated with tumour-associated tissues and skin cells. It can be used  
 CC to screen for antagonists and agonists, and to generate antibodies.  
 CC HKALL antagonists can be used to suppress excessive proteolysis and  
 CC subsequent skin cell scaling, so may be administered to treat skin  
 CC disorders. They may also be used to suppress excessive cell  
 CC proliferation, so can be administered to treat/prevent cancer, e.g.  
 CC adenocarcinoma, leukaemia and melanoma. Antibodies specific for HKALL  
 CC may be used directly as antagonists, or indirectly as a targeting or  
 CC delivery mechanism for bringing pharmaceutical agents to HKALL-expressing  
 CC cells. They are also useful to diagnose conditions/diseases characterised  
 CC by HKALL expression and to monitor therapeutic interventions. The  
 CC polynucleotide encoding HKALL, or complementary sequences, can be used to  
 CC produce hybridisation probes, useful to detect polynucleotides encoding  
 CC HKALL, e.g. to diagnose diseases relating to polypeptide expression  
 CC (e.g. cancers of the bladder, prostate) or monitor HKALL regulation  
 CC during therapeutic intervention. Polynucleotides encoding HKALL are  
 CC useful to produce antisense sequences for therapeutic administration to  
 CC modulate/prevent HKALL expression e.g. to treat/prevent skin disorders  
 CC or cancer a s above.  
 SQ Sequence 268 AA;

alignment\_scores:  
 Quality: 550.50 Length: 239  
 Ratio: 3.277 Gaps: 4  
 Percent Similarity: 70.293 Percent Identity: 45.607

alignment\_block:  
 us-09-030-606-173 x W94493 ..

Align seg 1/1 to: W94493 from: 1 to: 268  
 25 CAGCGGCACTGTCATGAA...AACGAATGTCCTGCGGCGCTCT 71  
 |||||  
 56 GlnAlaIleLeuLeuArgProHsnGlnLeuTyrCysGlyAlaValLe 72  
 72 GTGTCATCCGAGTGGTGTCTCAGCGGCACTGTTCCAGAACTCT 121  
 |||||  
 72 uValHisProGlnTrpLeuLeuThrAlaAlaHisCysArgLysValP 89

122 ACACCATGGGCTGGCCCTGCACAGTCTTGAGCGGCAAGACAGG 171  
 ::::|  
 89 heArgValArgLeuGlyHisTyrSerLeuSerProValTyrGlnSerGly 105  
 172 AGCCAGATGTGAGAGCCAGCCCTCCGTACGCGCACCAGATCAACAG 221  
 |||||  
 106 GlnGlnMetPheGlnGlyValTyrSerIleProHisProGlyTyrSerH 122  
 222 ACCCTGCTGCTACAGCCTCATGCTCATCACTTGAGCAGCAATCCGT 271  
 |||||  
 122 sProGlyHisSerAsnSplMetIleLeuTyrSleuAsnArgArgIleA 139  
 272 CCGAGTGTGACACCATCCGGAGCATGATGCTTGCGAGATGCTCACC 321  
 ::::|  
 139 rgrProHtLysAspValArgProIleAsnValSerSerHisCysProSer 155  
 322 GCGGGGAACTCTTCCTCGTTTGTGCTGGGGTCTGCTGCGCAAGGTGA 371  
 |||||  
 156 AlaGlyThrLysCysLeuValSerGlyTyrPely..... 166  
 372 CCTCAGCGGCTGTGTGTGCTCCCTCTTCAAGAGGCTCCTGCGCAATGCC 421  
 166 ..... 166  
 422 GGGGGCTGACCCAGAGCTGCTGCTCCAGCGAATGCTACGCTGCTGC 471  
 |||  
 167 .....ThrThrLysSerProGlnVal..HisPheProLysValLeuG 180  
 472 AGTCGCTGAACGTCTCGTGTGTGTGAGAGGCTGCTCAGTAACCTCTAT 521  
 |||||  
 180 IncysLeuAsnIleSerValLeuSerGlnLysArgCysGlyAlaPalatyr 196  
 522 GACCCGCTGTACACCCAGCATGTTCTGCGCGCGGAGGCGCAAGCA 571  
 |||||  
 197 ProArgGlnIleAspAspThrMetPheCysAla...GlyAspLysAlaG 212  
 572 GAAGGACCTCCGCAAGGCTGACTGGGGGCGCCCTGATGCAACGGGT 621  
 ::::|  
 212 YArgAspSerCysGlnGlyAspSerGlyGlyProValValCysAsnGlyS 229  
 622 ACTGTCAGGGGCTGTGTCTTCGAAAGCCCGCTGCGCAAGTGGC 671  
 |||||  
 229 erLeuGlnGlyLeuValSerTrpLysArgProCysAlaArgProAsn 245  
 672 GTGCCAGGTGTCTACACCACTCTGCAAAATTCAGTAGTGATAGAA 721  
 |||||  
 246 ArgProGlyValTyrThrAsnLeuCysLysPheThrLysTrpIleGln 262  
 722 AACGCTCAGGCCACT 737  
 ::::|  
 262 uThrIleGlnAlaAsn 267  
 seq\_name: A\_Geneseq\_36:R67888

seq\_documentation\_block:  
 ID R67888 standard; Protein: 253 AA.  
 AC R67888;  
 DT 09-AUG-1995 (first entry)  
 DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).  
 KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;  
 KM callostites; keratosis pilaris; ichthyoses; eczema.  
 OS Homo sapiens.  
 PN WO9500651-A.  
 PD 05-JAN-1995.  
 PF 20-JUN-1994; IB0166.  
 PR 18-JUN-1993; DK-000725.  
 PA (SYMB-) SYMBICOM AB.  
 PI Bgelrud T, Hansson L;  
 DR WPI: 95-052088/07.  
 DR N-PSDB: Q81203.  
 PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme  
 PT - and related vectors, transformed cells and polypeptides,

PT useful for treating skin disorders, e.g. acne or psoriasis, and  
 PT for identification of specific inhibitors.  
 CC Disclosure: Page 97: 137pp; English.  
 PS The enzyme encoded by this sequence is used in pharmaceutical, cosmetic  
 CC and skin care products, especially to treat and prevent acne,  
 CC xeroderma, or other hyperkeratotic conditions (e.g. callusities or  
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is  
 CC produced recombinantly following mammal, insect, plant, or  
 CC microorganism transformation with plasmid pss07.  
 SO Sequence 253 AA:

alignment\_scores: Quality: 519.50 Length: 247  
 Ratio: 3.038 Gaps: 5  
 Percent Similarity: 69.231 Percent Identity: 42.915

alignment\_block:  
 US-09-030-606-173 x R67888 ..

Align seg 1/1 to: R67888 from: 1 to: 253

```

1  GGCAGCCCG.....CACTCCAGCCCTGGCAGCGGCACTGGTCAT 41
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
33  G1yAlaProCysAlaArgGlySerHisProTprGlnValAlaLeuLeuSe 49
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
42  GGAAGAGCAATTTCTTCGTCGGGCGTCCTGGTACATCCGAGTGGGTCG 91
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
49  rGlyAsnGlnLeuHisCysGlyValLeuValAsnGlnGluArgTyrVal 66
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
92  TGTGAGCCGCACTGTTTCCAGAACTCTACACCACTGGGCGGCTGG 141
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
66  eutHraIdAlaHisCysLysMetAsnGlnuYrThrValHisLeuGlySer 82
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
142  CACAGTCTTGAGGCGCCAGCAAGCCAGGAGCCAGATGGGAGGCCAG 191
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
83  AspThrLeu...GlyAspArgArg.....AlaGlnArgIleLysAlaSe 96
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
192  CCTTCGCTGAGCGCAGCCAGATACAGACACCTTCTCGCTAGAGACC 241
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
96  rLySerPheArgHisProGlyTyrSerThrGlnThrHisValAsnAsp 113
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
242  TCATGCTCATCACTGAGCAATCCGTGTCCAGTGTGACACCATCCGG 291
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
113  euMetLeuValLysLeuAsnSerGlnAlaArgLeuSerMetValLys 129
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
292  AGCATCAAGCATGCTGTCCAGATGCCCTACCGGGGAACCTTCTCGT 341
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
130  LyValAlaArgLeuProSerArgCysGluProGlyThrThrCysThrVa 146
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
342  TTTCAGCTGGGCTGCTGGCAGACGGTGAAGCTCAGGGTGTGTCTGC 391
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
146  lSerGlyTyrProLy..... 150
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
392  CCTTCTCAAGAGGCTCTGTCCAGTCCGCGGGGCTGACCCAGACGCT 441
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
151  .....ThrThrThrSer 154
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
442  GCGTCCAGGAGCAATGCTACCTGCTGCAAGTCCGTGAACGTGCGTG 491
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
155  ProMetValThr..PheProSerAspLeuMetCysValAspValLysLeu 170
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
492  GTGTGAGAGAGTGTGAGTACAGTATGACCGGCTGACACCCAG 541
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
171  lIleSerProGlnAspCysThrLysValLysAspLeuLeuGlnAsnSe 187
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
542  CATGTTTGGCGCCGCGAGGCGAGCAAGACCAAGAGACTCTGCAACGCTG 591
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
187  rMetLeuCysAlaGlyLysIleProAspSerLysLysAsnAlaCysAsnGly 204
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
592  ACTTGGGGGGCCCTGATCTGCAAGGAGTACTTGCAGAGGCTTGTGTCT 641
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
204  spSerGlyGlyProLeuValCysArgGlyThrThrLeuGlnGlyLeuValSer 220

```

642 TTCGAAAAGCCCGCTGTGGCAAGTTGGGCTGCAGGTGTACACAA 691  
 221 TrGclYThrPheProCysGlyGlnProAsnAspProGlyValIlyThrGcl 237  
 692 CCTTCGCAATTCATGAGTGTAGTATGAGAAAACCTTCAG 731  
 237 nValCysLysPheThrLysTrpLysAsnAspThrMetLys 250

seq\_name: A\_Geneseq\_36:W05383

seq\_documentation\_block:

ID W05383 standard; Protein: 253 AA.  
 AC W05383;  
 DT 31-DEC-1996 (first entry)  
 DE Human amyloid precursor protein protease.  
 KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;  
 KW therapy.  
 OS Homo sapiens.  
 PN W09631122-A1.  
 PD 10-OCT-1996.  
 PF 02-APR-1996; U04294.  
 PR 04-APR-1995; US-416257.  
 PA (ELIL ) LILLY & CO B.LI.  
 PI Dixon EP, Johnstone EM, Little SP;  
 DR WPI: 96-464694/46.  
 DR N-PSDB: T39783.  
 PT New isolated human amyloid precursor protein protease - used to  
 PT develop prods. for the treatment or diagnosis of associated  
 PT conditions, esp. Alzheimer's disease  
 PS Claim 1; Page 44-45; 55pp; English.  
 CC Human amyloid precursor protein protease (W05383) is involved in  
 CC the processing or clearance of amyloid precursor protein to form  
 CC beta-amyloid peptide. Its amino acid sequence was deduced from  
 CC a cDNA clone (T39783) obtd. from a human lung library. Recombinant  
 CC protease can be produced in transformed or transfected prokaryotic  
 CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is  
 CC used to develop products for the design and testing of cpds. useful  
 CC for treating or preventing conditions associated with beta-amyloid  
 CC peptide, esp. Alzheimer's disease.  
 SO Sequence 253 AA:

alignment\_scores: Quality: 519.50 Length: 247  
 Ratio: 3.038 Gaps: 5  
 Percent Similarity: 69.231 Percent Identity: 42.915

alignment\_block:  
 US-09-030-606-173 x W05383 ..

Align seg 1/1 to: W05383 from: 1 to: 253

```

1  GGCAGCCCG.....CACTCCAGCCCTGGCAGCGGCACTGGTCAT 41
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
33  G1yAlaProCysAlaArgGlySerHisProTprGlnValAlaLeuLeuSe 49
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
42  GGAAGAGCAATTTCTTCGTCGGGCGTCCTGGTACATCCGAGTGGGTCG 91
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
49  rGlyAsnGlnLeuHisCysGlyValLeuValAsnGlnGluArgTyrVal 66
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
92  TGTGAGCCGCACTGTTTCCAGAACTCTACACCACTGGGCGGCTGG 141
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
66  eutHraIdAlaHisCysLysMetAsnGlnuYrThrValHisLeuGlySer 82
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
142  CACAGTCTTGAGGCGCCAGCAAGCCAGGAGCCAGATGGTGGAGCCAG 191
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
83  AspThrLeu...GlyAspArgArg.....AlaGlnArgIleLysAlaSe 96
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
192  CCTTCGCTGAGCGCAGCCAGATACAGACACCTTCTCGCTAGAGACC 241
   |||:||||| |||:|||||:|||||:|||||:|||||:|||||:
96  rLySerPheArgHisProGlyTyrSerThrGlnThrHisValAsnAsp 113

```



AC W12393: 15-MAY-1997 (first entry)  
 DT Mouse neuropsin protein.  
 KW Mouse: neuropsin; hippocampus; lambda gt10; primer; PCR; amplification;  
 KW polymerase chain reaction; serine protease domain; nerve growth factor;  
 KW NGF; insect cell; virus; expression vector; transfection;  
 KW cerebral disease.  
 OS Mus musculus.  
 PN J08311099-A.  
 PD 26-NOV-1996.  
 PF 13-MAR-1996: 056367.  
 PR 14-MAR-1995: JP-054584.  
 PA (SHIO/) SHIOZAKA S.  
 DR WPI: 97-061812/06.  
 DR N-PSDB: T63251.  
 PT Nucleic acid encoding neuropsin - for producing neuropsin, useful  
 for diagnosis and treatment of cerebral disease  
 PS Claim 1, Page 6-7, 9pp; Japanese.  
 CC This is the amino acid sequence of a novel mouse protein designated  
 CC neuropsin. The encoding gene was isolated from a mouse hippocampal  
 CC cDNA library in lambda gt10 using a cloned, amplified fragment of  
 CC the gene (clone B41; T63254). This fragment was amplified using  
 CC primers T63252-3. The primers were synthesised based on the serine  
 CC protease domain of nerve growth factor (NGF)-gamma. The screen isolated  
 CC 6 positive clones, of which clone NP5 contained the longest insert  
 CC (this sequence). The protein has a molecular weight of around 26 kD.  
 CC It has 43% homology with EGF-BP, 41 % with NGF-gamma; 39% with NGF-alpha;  
 CC 38% with trypsin and 18% with tPA. The protein can be used for clinical  
 CC diagnosis and treatment of cerebral diseases.  
 SQ Sequence 260 AA;

alignment\_scores:  
 Quality: 496.00 Length: 244  
 Ratio: 2.851 Gaps: 5  
 Percent Similarity: 71.311 Percent Identity: 38.934

alignment\_block:  
 US-09-030-606-173 x W12393 ..

Align seg 1/1 to: W12393 from: 1 to: 260

```

7  CCGACATCGACGCGCTGCGAGCGCGACTGTCATGGAACGAATTGTT 56
  |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
41  ProHisserGlnProTyrGlnAlaAlaLeuPheGlnGlyGluAValleu 57
57  CTGCTCGGCGCTCTGCTGCTATCGCGAGTGGTGTCTGCAAGCCGCAACT 106
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
57  eCysGlyGlyValleuValGlyAspArgTyrValleuThrAlaAlaHisC 74
107  GTTTCAGACTCCTACACATCGGCTGCGGCTGACACAGTCTGAGGCC 156
  ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
74  yslYslYslGlnLysTyrSerValArgleuGlyAspHisSerleuGlnSer 90
157  GACCAAGACCGAGGAGCCAGATGTGTGAGGCGCAGCTCTCCGATGCGCA 206
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
91  ArgAspGlnPro...GlnGlnGlnIleGlnValAlaGlnSerIleGlnH1 106
207  CCCGAGTATACACAGACCCCTTG.....CTGCGTACAGACTCATGC 247
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
106  sProCysTyrAsnAsnSerAsnProGluAspHisSerHisAspIleMetL 123
248  TCATCAGATTGAGCAATCCGTCGCGAGTGTGACACCATCGGAGAGCAT 297
  |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
123  euIleAglleuGlnAsnSerAlaAsnleuGlyAspLysValLysProVal 139
298  AGCATTCCTTCGACGTCCCTACCGCGGGAACACTTGTCCCTGTTGCG 347
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
140  GlnleuAlaAsnleuCysProLysValGlyGlnLysCysIleIleSerG1 156
348  CTGGGGTCTGCTGGCGAACGGTGAAGCTCAGCGGCTGTGTCTGCCCTTT 397
  |||||||
156  yTyrGly..... 158

```

```

398  CAAGAGAGTCTCTGCGCCAGTCCGGGGGCTGACCCAGAGCTGTGGCTCC 447
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
159  .....ThyValThrSerProGln 164
448  CAGGACAGATGCTTACCGTCTGCTGCTGCGTGAACGTGTGGTGTGCTG 497
  :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
165  Glu..AsnProAsnThrLeuAsnCysAlaGlnValLysIleTyrSer 180
498  GAGAGAGTCTGCACTAGTACTATGACCCGCTGTACCACCCAGCATGTT 547
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181  GlnAsnLysCysGlnValArgAlaTyrProGlyLysIleThrGlnLysVal 197
548  CTGCGCGCGCGAGGCGCAAGACGACTCTCTGCTGACAGGTGACTCTG 597
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
197  LysAla..GlySerSerAsnLysAlaAspThrCysGlnGlyAspSerG 213
213  LysLysProLeuValLysAspGlyMetleuGlnLysIleThrSerTyrGly 229
648  AAGACCCCTGCTGCTGCAAGTGGCGTCCAGGTGTCTACACCACTCTG 697
  :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
230  SerAspProCysGlyLysProGluLysProGlyValTyrThrLysIleC 246
698  CAATTCACCTAGTGTAGAGAAACCGTC 728
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
246  sArgTyrThrThrTyrIleLysIleThrMet 256

```

seq\_name: A\_Geneseq\_36:W87703

seq\_documentation\_block:

ID W87703 standard; Protein; 260 AA.

```

AC W87703:
DT 09-MAR-1999 (first entry)
DE A human serine protease designated HGBAB90.
KW Human serine protease; HGBAB90; pulmonary emphysema; pancreatitis;
KW arthritis; multiple sclerosis; peridontal disease; cystic fibrosis;
KW respiratory disease; thrombosis; cancer; cachexia; angina; glaucoma;
KW inflammatory disorder; osteoporosis; cardiovascular disorder;
KW hypertension; atherosclerosis disorder; cardiac infarction;
KW stroke; asthma; psoriasis; chronic neurodegenerative disease;
KW Alzheimer's; Parkinson's; Huntington's; demyelinating disease;
KW immune deficiency; photoreceptor degeneration; lens cataract formation;
KW organ transplant rejection; cataract; retinosis; muscular dystrophy;
KW renal failure; cerebral vasospasm; diabetic nephropathy.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 56 /note="encoded by YTA"
FT ID EP-887414-A2.
PD 30-DEC-1998.
PF 21-APR-1998: 303064.
PR 01-DEC-1997: EP-309646.
PR 09-JUN-1997: GB-011952.
PA (SMK) SMITHKLINE BEECHAM PLC.
PI Burgess NA, Clinkenbeard HE, Southan CD;
DR WPI: 99-047881/05.
DR N-PSDB: V84052.
PT New human serine protease HGBAB90 polypeptide and polynucleotide -
PT useful as diagnostic reagents and for prevention and treatment of
PT cancer, Alzheimer's disease and AIDS immune deficiency
PS Claim 1, Page 13, 19pp; English.
CC The present sequence represents a human serine protease (designated
CC HGBAB90). HGBAB90 polypeptides and polynucleotides are useful for
CC diagnosing susceptibility to diseases by detecting mutations in the
CC HGBAB90 gene, and can diagnose diseases associated with HGBAB90 imbalance
CC by determining HGBAB90 polypeptide expression levels. HGBAB90
CC polypeptides can be used to screen for agonists and antagonists,
CC which can be used in treatment to activate (agonist) or inhibit
CC (antagonist) HGBAB90 activity, in addition to direct administration of
CC antisense sequences or soluble HGBAB90 polypeptides which competitively
CC bind HGBAB90 ligands to prevent expression. Gene therapy may also be used

```

CC to affect endogenous HGBAB90 polypeptide production. HGBAB90 antibodies  
 CC are useful for inducing an immune response to immunise and prevent  
 CC diseases. Diseases that can be treated are pulmonary emphysema,  
 CC arthritis, multiple sclerosis, periodontal disease, cystic fibrosis,  
 CC respiratory disease, thrombosis, cancer, cachexia, angina, glaucoma,  
 CC inflammatory disorders, osteoporosis, cardiovascular disorders such as  
 CC hypertension, atherosclerosis disorders such as cardiac infarction and  
 CC stroke, asthma, psoriasis, chronic neurodegenerative diseases such as  
 CC Alzheimer's, Parkinson's and Huntington's, demyelinating diseases, AIDS  
 CC immune deficiency, disorders of photoreceptor degeneration and lens  
 CC cataract formation, organ transplant rejection, cataracts, restenosis,  
 CC muscular dystrophy, renal failure, cerebral vasospasm, pancreatitis and  
 CC diabetic neuropathy.  
 SO Sequence 260 AA:

alignment\_scores:  
 Quality: 459.00 Length: 242  
 Ratio: 2.905 Gaps: 5  
 Percent Similarity: 65.289 Percent Identity: 39.256

alignment\_block:  
 US-09-030-606-173 x W87703 ..

Align seg 1/1 to: W87703 from: 1 to: 260

```

7  CCGGACCTCGAGCCCTGGCAGCGGCGGCTGTCATGGAAGAAATGTT 56
  |||||||||||||||||||||||||||||||||||||||||||||||
41  ProHisSerGlnProTrrpGlnAlaAlaLeuPheGlnGlnGlnLeu 57
  |||||||||||||||||||||||||||||||||||||||||||||||
57  CTGCTCGGCGCTCTGTGTCATCCGCGAGTGGTGTCTACCGCGACACT 106
  |||||||||||||||||||||||||||||||||||||||||||||||
57  ucyeGlyValLeuValGlyGlyAsnTrrpValLeuThrAlaAlaHis 74
  |||||||||||||||||||||||||||||||||||||||||||||||
107  GTTCCAGACTCTCTACACCATCGGCGCTGGCGCTGCACAGTCTTGAGCC 156
  |||||||||||||||||||||||||||||||||||||||||||||||
74  yslYslYslProLysTrrpThrValArgLeuGlnGlyAspHisSerLeuGlnAsn 90
  |||||||||||||||||||||||||||||||||||||||||||||||
157  GACCAAGAGCCGAGGAGCCAGATGGTGGAGGCGAGCCCTCCGTCAGCA 206
  |||||||||||||||||||||||||||||||||||||||||||||||
91  LysAspGlyPro...GluGlnGlnLeuProValValGlnSerIleProH 106
  |||||||||||||||||||||||||||||||||||||||||||||||
207  CCCAGAGTACAACAGACCCCTTGTCTGCT.....AAGCACTCATGC 247
  |||||||||||||||||||||||||||||||||||||||||||||||
106  sProCysTrrpAsnSerSerAspValGlnAspHisAsnHisAspLeuMet 123
  |||||||||||||||||||||||||||||||||||||||||||||||
248  TCATCAAGTGGAGCAATCCGTGTCGAGTGTACACCATCGGAGCATC 297
  |||||||||||||||||||||||||||||||||||||||||||||||
123  eulGlnLeuArgAspGlnAlaSerLeuGlnGlySerLysValLysProIle 139
  |||||||||||||||||||||||||||||||||||||||||||||||
298  AGCATTCGTTGGCAGTGGCTTACCGCGGGAACCTCTGCTCGTTCTGG 347
  |||||||||||||||||||||||||||||||||||||||||||||||
140  SerLeuAlaAspHisCysThrGlnProGlnLysCysThrValSerGln 156
  |||||||||||||||||||||||||||||||||||||||||||||||
348  CTGGGGTCTGCTGGGCAAGGAGTCAACGGGTGTGTCTGCCCTCTT 397
  |||||||
156  yTrrpGly..... 158
  |||||||
398  CAAGAGGCTCTGTGCCAGTGGCGGGGCTGACCCAGAGCTTCGCTCC 447
  |||||
159  .....ThrValThrSerProArg 164
  |||||
448  CAGGAGAAATGCTACGCTGCTGACGCTGCGAAGCTGCGGTGCTCT 497
  |||||||
165  Glu...AsnPheProAspThrLeuAsnCysAlaGlnValLysIlePhePro 180
  |||||||
498  GAGGAGTCTGCTCAGTAACTCTATGACCGGCTGTACACACCCAGCATGT 547
  |||||||
181  GlnYslYslCysGlnAspAlaTrrpProGlnLysIleThrAspGlyMetVal 197
  |||||||
548  CTCGGCGCGCGGAGGCAAGACCAAGAGACTCTGCAACGGTCACTGCT 597
  |||||||
197  LcysAlaGlnSerSerLysGly...AlaAspThrCysGlnGlnLysAspSerG 213
  |||||||

```

```

598  GGGGCCCCGATGTCGACACGGGTACTTGCAGGCGCTTGTGTCTTCGA 647
  |||||||||||||||||||||||||||||||||||||||||||||||
213  LysGlnProLeuValCysAspGlyAlaLeuGlnGlnYlleThrSerTrrpGly 229
  |||||||||||||||||||||||||||||||||||||||||||||||
648  AAAGCCCCGTTGGCCAGTGGCGGCTGCAGAGTGTACACCAACCTCTG 697
  |||||||
230  SerAspProCysGlyArgSerAspLysProGlnValYrrThrAsnIleCys 246
  |||||||
698  CAATTCACTCAGTCGATAGAGAA 722
  |||||||
246  SArgTrrLeuAspTrrIleLysLys 254
  |||||||

```

seq\_name: A\_Geneseq\_36:W45395

seq\_documentation\_block:  
 ID W45395 standard; Protein: 237 AA.  
 AC W45395;  
 DT 06-JUL-1998 (first entry)  
 DE Mature prostate-specific glandular kallikrein hK2.  
 KW Prostate-specific glandular kallikrein: hK2 protein; antigen;  
 KW prostate carcinoma; prostate cancer; benign prostatic hyperplasia;  
 KW diagnosis; human.  
 OS Homo sapiens.  
 PN W09802748-A1.  
 PD 22-JAN-1998.  
 PF 15-JUL-1997: U12322.  
 PA (HYBR-) HYBRITRICH INC.  
 PA (MAYO-) MAYO FOUNDATION.  
 PI Grauer L, Klee GG, Mikolajczyk SD, Saeed M, Tindall DJ,  
 PI Young CYF;  
 DR WPI: 98-120378/11.  
 DR N-PSDB: V06602.  
 PT Diagnostic methods using antibodies which bind prostate antigens -  
 PT useful for, e.g. monitoring treatment or progression of prostate  
 PT cancer  
 PS Example 1: Page 49-50: 100pp: English.  
 CC This polypeptide comprises human mature prostate-specific glandular  
 CC kallikrein hK2. It is encoded by cDNA (see V06602) derived from  
 CC human benign prostate hyperplasia (BPH) tissue RNA. The invention  
 CC provides a diagnostic method comprising contacting antibodies  
 CC that specifically bind to pro-hK2 (pHK2, see W45396) or mature hK2  
 CC with a sample of physiological fluid from a human. The assay  
 CC is based on the discovery that pHK2 is detected in the  
 CC supernatant of a prostate carcinoma cell line and that hK2 is  
 CC present in human physiological fluid from prostate cancer cells.  
 CC The method is useful for monitoring the treatment and/or  
 CC progression of prostate cancer, or for the early detection of  
 CC prostate cancer in males that have BPH or a high grade prostatic  
 CC neoplasia (HPN) or whose family members have or had BPH, HPN or  
 CC a prostate cancer. Monitoring the presence and/or amount of hK2  
 CC complexes with plasma proteins may also be important in  
 CC distinguishing between prostate cancer and BPH.  
 SO Sequence 237 AA:

alignment\_scores:  
 Quality: 449.50 Length: 254  
 Ratio: 2.863 Gaps: 5  
 Percent Similarity: 61.811 Percent Identity: 37.402

alignment\_block:  
 US-09-030-606-173 x W45395 ..

Align seg 1/1 to: W45395 from: 1 to: 237

```

10  CACTCGAGCCCTGGCAGGCGGCACTGTATGGAAGAAATGTTCTG 59
  |||||||||||||||||||||||||||||||||||||||||||||||
10  HisSerGlnProTrrpGlnValAlaValLysSerHisGlyTrrpAlaHisCys 26
  |||||||
60  CTCGGCGCTCTGCTGTCATCCGAGTGGGTGCTGTACGCGCACACTGTT 109
  |||||||

```

```
26  sglYgLYvalleuValHisProGlnTrpValIleuThrAlaAlaHisCysL 43
110  TCCGAGACTCTTACACATCGGGCTGGCCCTGCACAGTCTTGAGCGGAC 159
    :::::::::::::: :::: ||||| ||::::::::||
43  euLYsLYsAsnSerGlnValIrpLeuGLYArGHisAsnLeuPhGluPro 59
160  CAAGAGCCAGGAGACAGATGCTGAGAGCCAGCCTCTCCGTACGGCACCC 209
    ::::: ||| ||| :::: ||| |||
60  GlnAspThrGly...GlnArgValProValSerHisSerPheProHisPr 75
210  AGAGTACACAGACCTTGCTC..... 231
    | ||||| |||||
75  oLeuTYrAsnMetSerLeuLeuLYsHisGlnSerLeuArgProAspGluA 92
232  ....GCTAAGACCTCATGCTCATCAAGTATGACGATCCGTCGAG 276
    ::||::||::||::||::||::||::||::||::||::||::||::||
92  sPserSerHisAspLeuMetLeuLeuArgLeuSerGluProAlaLYsIle 108
277  TCTGACACATCCGAGACATCAGCATTCCTGCGAGTGCCCTACCGCGG 326
    ::||::||::||::||::||::||::||::||::||::||::||
109  ThrAspValLYsValLeuGLYLeuProThrGlnGluProAlaLeuGL 125
327  GAACTCTTCCTCTTCTGCTGGCTGGGCTGCTGCGACGAGTGAGCTCA 376
    |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
125  YThrThrCysTYrAlaSerGlyTrpGlySerIle..Glu..... 137
377  CGGGTGtGTGTCGCCCTTCAAGAGAGTCTGCGCCAGTCCGCGGGG 426
137  ..... 137
427  CTGACCCAGAGCTCTGCGTCCAGGAGAAATCCCTACCGTCTGACGTGC 476
    ||||| :::: ||| |||||
138  ....ProGlnGluPhen.....ArgProArgSerLeuGlnCys 149
477  GTGACGTGTCGGTGTCTGAGAGAGTCTGCAGTAAGCTATGACCC 526
    ||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
150  ValSerLeuHisLeuLeuSerAsnAspMetCysAlaArgAlaTYrSerGI 166
527  GCTGTACCACCCCGACATGTTCTGCGCGCGGCGAGGCGAAGACCAGAAG 576
    ||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
166  uLYsValIThrGluPhenMetLeuCYsAlaGLYLeuTrpThrGLYGLYLYsA 183
577  ACTCTGCAACGGTACTGTGGGGGCCCTGATCTGCAACGGGTACTTG 626
    ||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
183  sPThrCysGLYGLYAspSerGLYGLYProLeuValCYsAsnGLYValLeu 199
627  CAGGGCCTTGCTTTCGGAAGACCCCGTGTGCCCAAGTTGGCTGCC 676
    |||||:::|::|::|::|::|::|::|::|::|::|::|::|::|
200  GlnGLYIleThrSerTrpGlyProGluProCYsAlaLeuProGluLYsPr 216
677  AGGTGTACACCACTCTGCAAAATTCAGTGTAGTAGAGAAACCG 726
    ||:::|::|::|::|::|::|::|::|::|::|::|::|::|::|
216  oAlaValIYThrLYsValValHisTYrArgLYsTrpIleLYsAspThri 233
727  TCCAGGCCAGT 737
    :: ||:::
233  leAlaAlaAsn 236
```



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:23:29 ; Search time 2928.69 seconds  
(without alignments)  
1373.683 Million cell updates/sec

Title: US-09-030-606-173  
Perfect score: 1265  
Sequence: 1 GGCAGCCCGCCACTCGCAGCC.....AGAGANGGCCAAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: \*  
1: gb\_dal: \*  
2: gb\_dal: \*  
3: gb\_dal: \*  
4: gb\_dal: \*  
5: gb\_dal: \*  
6: gb\_dal: \*  
7: gb\_dal: \*  
8: gb\_dal: \*  
9: gb\_dal: \*  
10: gb\_dal: \*  
11: gb\_dal: \*  
12: gb\_dal: \*  
13: gb\_dal: \*  
14: gb\_dal: \*  
15: gb\_dal: \*  
16: gb\_dal: \*  
17: gb\_dal: \*  
18: gb\_dal: \*  
19: gb\_dal: \*  
20: gb\_dal: \*  
21: gb\_dal: \*  
22: gb\_dal: \*  
23: gb\_dal: \*  
24: gb\_dal: \*  
25: gb\_dal: \*  
26: gb\_dal: \*  
27: gb\_dal: \*  
28: gb\_dal: \*  
29: gb\_dal: \*  
30: gb\_dal: \*  
31: gb\_dal: \*  
32: gb\_dal: \*  
33: gb\_dal: \*  
34: gb\_dal: \*  
35: gb\_dal: \*  
36: gb\_dal: \*  
37: gb\_dal: \*  
38: gb\_dal: \*  
39: gb\_dal: \*  
40: gb\_dal: \*  
41: gb\_dal: \*  
42: gb\_dal: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	972.4	76.9	1347	42	AF113140 Homo sapi

2	553	43.7	5900	42	AF113141	AF113141 Homo sapi
3	474.8	37.5	4740	42	AF135023	AF135023 Homo sapi
4	474.8	37.5	4385	42	AF148532	AF148532 Homo sapi
5	452	35.7	1140	3	SSU76256	U76256 Sus scrofa
6	310	24.5	1237	12	AF019979	AF019979 Mus muscu
7	205.4	16.2	35197	11	AC005782	AC005782 Homo sapi
8	190.8	15.0	6592	10	HSCPT1T5	X90780 Homo sapien
9	190.2	15.0	37785	9	HUMXRCC1G	L34079 Human XRCC1
10	184.6	14.6	8174	5	I34189	I34189 Sequence 5
11	174.2	13.8	281	10	HUMRSSA19	K03500 Human 37 bp
12	169	13.4	1655	9	D87943	D87943 Human DNA f
13	167.4	13.2	205251	11	AC005392	AC005392 Homo sapi
14	165.4	13.1	6532	10	HSCPT1T5	X90780 Homo sapien
15	163.8	12.9	4033	9	AB006136	AB006136 Homo sapi
16	163.8	12.9	42710	42	AC006942	AC006942 Homo sapi
17	162.2	12.8	16689	10	HSN011712	AC0011712 Homo sapi
18	160.2	12.7	16689	10	HSN011712	AC0011712 Homo sapi
19	158.8	12.6	35465	11	AC005525	AC005525 Homo sapi
20	145.8	11.5	357	9	HUMSATVIS	M54997 Human DNA v
21	145.2	11.5	4067	5	I30049	I30049 Sequence 1
22	145.2	11.5	4067	10	S51329	S51329 AAVS1-aden
23	144.4	11.4	37314	9	HUMAMDA	M63786 Human DNA f
24	140.6	11.1	4776	11	HSBCL352	U05681 Human prote
25	134.4	10.6	111084	11	AC006486	AC006486 Homo sapi
26	132.2	10.5	2403	10	HSN011712	U31342 Human nucle
27	129	10.2	3840	10	HSAPOC2G	X05151 Human apoc-
28	125.8	9.9	4340	10	HUMAPOC1	M10612 Human apol1
29	113.2	8.9	4077	10	HUMAPOC2	J02698 Human apol1
30	112	8.9	4776	11	HSBCL352	U05681 Human prote
31	111.2	8.8	2711	9	AB012761	AB012761 Homo sapi
32	110.2	8.7	6500	42	AF095743	AF095743 Homo sapi
33	109.8	8.7	1438	5	E13202	E13202 Human gene
34	109.8	8.7	1419	9	D78203	D78203 Homo sapien
35	109.8	8.7	1506	10	HS062801	U62801 Human prote
36	109.8	8.7	1451	11	AF013988	AF013988 Homo sapi
37	106.2	8.4	732	5	I95869	I95869 Sequence 2
38	102	8.1	5574	11	AF055481	AF055481 Homo sapi
39	97.4	7.7	586	14	G41906	G41906 SHGC-56840
40	95	7.5	986	5	A42048	A42048 Sequence 1
41	95	7.5	969	11	HUMSERPROT	L33404 Human strat
42	90	7.1	881	12	RNTRYPVB	X59013 Rat mRNA fo
43	89.6	7.1	729	11	HSN011712	U92322 Homo sapien
44	89.2	7.1	893	4	PTTRYP	X56744 P. platessa
45	88.8	7.0	877	4	PMTRYPIN	X82223 P. magellani

## ALIGNMENTS

RESULT 1	AF113140	1347 bp	MRNA	PRI	25-MAR-1999
LOCUS	AF113140				
DEFINITION	Homo sapiens serine protease mRNA, complete cds.				
ACCESSION	AF113140				
NID	94512029				
VERSION	AF113140.1	GI:4512029			
KEYWORDS	human.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
AUTHORS	1 (bases 1 to 1347)				
TITLE	Wang, K.				
JOURNAL	Molecular cloning and characterization of prostate, an androgen-regulated serine protease with prostate-restricted expression				
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)				
REFERENCE	99179024				
AUTHORS	2 (bases 1 to 1347)				
TITLE	Nelson, P.S., Gan, L., Ferguson, C., Moss, P., Gellinas, R., Hood, L. and Wang, K.				
	Direct Submission				

JOURNAL Submitted (10-DEC-1998) Chiroscience R & D, 1631 220th St SE,  
Bothell, WA 98021, USA  
FEATURES Location/Qualifiers  
source 1. 1347  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13"  
1. 765  
/note="androgen regulated"  
/codon\_start=1  
/product="serine protease prostase"  
/protein\_id="AAD21580.1"  
/db\_xref="PID:94512030"  
/db\_xref="GI:4512030"  
/translation="MATAGNPMGMFLGYLLIIVAGSLVSGSCQIINSEHCSPHSQPW  
QAAIWMENLEFCSGVYVHPQWYLSAHCNOSTTICLGLHSEADPEPSQVYASLS  
VRIPEYNPDLANDLMLIKLDESVSSTIRISISQCPYAGNSCLVSGWLLANGR  
MPVLQCAVSVSEVCSKLYDPLVHPMFCAGGHDGKDSQNGDSGLIIONGYLQ  
GLVSEKAPCGGVGVYVYTNLCKETEMETKVOAS"  
1327..1332

polya\_site 269 a 489 c 334 g 255 t  
BASE COUNT  
ORIGIN

Query Match 76.9%; Score 972.4; DB 42; Length 1347;  
Best Local Similarity 87.0%; Pred. No. 3.5e-214;  
Matches 1149; Conservative 0; Mismatches 22; Indels 150; Gaps 2;

QY 2 GCAGCCGCACTGCGACCCCTGGCAGGCGCAGCTGTCATGAAACGAATTGTTGCT 61  
DB 110 GCAGCCGCACTGCGACCCCTGGCAGGCGCAGCTGTCATGAAACGAATTGTTGCT 169  
QY 62 CGGCGCTCGTGGCATCGCAGTGGTGTGTCAGCCGACACCTGTTCCAGAACCT 121  
DB 170 CGGCGCTCGTGGCATCGCAGTGGTGTGTCAGCCGACACCTGTTCCAGAACCT 229  
QY 122 ACACCATCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGG 181  
DB 230 ACACCATCGGCTGGGCTGCACAGTCTTGAAGCCGACCAAGAGCCAGGAGCCAGATGG 289  
QY 182 TGAAGCCACCTCTCTCCGTACGCGACCAAGTACACAGACCTTGTCTGCTACAGAC 241  
DB 290 TGAAGCCACCTCTCTCCGTACGCGACCAAGTACACAGACCTTGTCTGCTACAGAC 349  
QY 242 TCATGCTCATCAAGTTGAGAGAAATCGGTCCGAGTGCAGACATCCGAGATCGCA 301  
DB 350 TCATGCTCATCAAGTTGAGAGAAATCGGTCCGAGTGCAGACATCCGAGATCGCA 409  
QY 302 TTGCTTGCAGTGCCTTACCGCGGGAACCTTGCTGTTTGTGCTGGGCTGCTGG 361  
DB 410 TTGCTTGCAGTGCCTTACCGCGGGAACCTTGCTGTTTGTGCTGGGCTGCTGG 469  
QY 362 CGAACGGTGAGCTTCACGGGTGTGTGTCTTCCCTTTCAGAGAGTCTTGGCCAGTGC 421  
DB 470 CGAAC----- 475  
QY 422 GGGGGGTGACCAAGAGTCTGCGTCCAGGAGAAATGCTACCGTGTGAGTGGCGTAA 481  
DB 475 -----GGCAGAAATGCTACCGTGTGAGTGGCGTAA 506  
QY 482 CGTGTGCGTGTGTCTGAGAGAGTCTGACATAGCTTATGACCCGCTGTACACCCAG 541  
DB 507 CGTGTGCGTGTGTCTGAGAGAGTCTGACATAGCTTATGACCCGCTGTACACCCAG 566  
QY 542 CATGTTCTGCGCGCGGAGGAGCAAGACCAAGAGACTCTTGCACCGGTACTCTGGGG 601  
DB 567 CATGTTCTGCGCGCGGAGGAGCAAGACCAAGAGACTCTTGCACCGGTACTCTGGGG 626  
QY 602 GCCCGTATCTGCAAGGGGACTTGTGAGGGGCTGTGCTTTTGGGAAACCCCGTGTG 661  
DB 627 GCCCGTATCTGCAAGGGGACTTGTGAGGGGCTGTGCTTTTGGGAAACCCCGTGTG 686

QY 662 CCAAGTTGGCGTGCCAGTGTCTACACCAACCTTGCAATTCAGTGTGATAGAGA 721  
DB 687 CCAAGTTGGCGTGCCAGTGTCTACACCAACCTTGCAATTCAGTGTGATAGAGA 746  
QY 722 AACCTTCAGGCGCAGTTAACTCTGGGAGTGGGAACCAATGAATTTGACCCCAATACA 781  
DB 747 AACCTTCAGGCGCAGTTAACTCTGGGAGTGGGAACCAATGAATTTGACCCCAATACA 806  
QY 782 TCCGTGGGAAGGAATTCAGGAATATCTGTTCCAGGCCCTCTCCCTCAGGCCAGGAGT 841  
DB 807 TCCGTGGGAAGGAATTCAGGAATATCTGTTCCAGGCCCTCTCCCTCAGGCCAGGAGT 866  
QY 842 CCAGGCCCGCCAGGCCCTCTCCCTCAACCAAGGATACAGATCC----- 886  
DB 867 CCAGGCCCGCCAGGCCCTCTCCCTCAACCAAGGATACAGATCCCGAGCCCTCTCTCT 926  
QY 886 -----CCAGCCCT 894  
DB 927 CAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCCAGCCCT 986  
QY 895 CCTCCCTCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCC 954  
DB 987 CTTCCCTCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCC 1046  
QY 955 AGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCC 1014  
DB 1047 AGCCCTCTCTCTCTCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCC 1106  
QY 1015 GGGGTGAGGCCCGCCAGCCCTCTCTCTCAGAGTACAGAGTCCAGACCCCGCCAGT 1074  
DB 1107 GGGGTGAGGCCCGCCAGCCCTCTCTCTCAGAGTACAGAGTCCAGAGGCCCGCCAGT 1166  
QY 1075 GTTCCCGCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCC 1134  
DB 1167 CTTCCCGCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCC 1226  
QY 1135 GCCACCTAGATTTTCCCTTACACAGTCCGCCCTTGTGANGATTGACCAACTTACCA 1194  
DB 1227 GCCACCTAGATTTTCCCTTACACAGTCCGCCCTTGTGANGATTGACCAACTTACCA 1286  
QY 1195 GTTGTGTTTCAATTTTNGTCCCTTCCCTTGTGANGATTGACCAACTTACCA 1254  
DB 1287 GTTGTGTTTCAATTTTNGTCCCTTCCCTTGTGANGATTGACCAACTTACCA 1346  
QY 1255 C 1255  
DB 1347 C 1347

RESULT 2  
AF113141  
LOCUS AF113141 5900 bp DNA PRI 25-MAR-1999  
DEFINITION Homo sapiens serine protease prostase gene, complete cds.  
ACCESSION AF113141  
NID 94512031  
VERSION AF113141.1 GI:4512031  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and  
Wang,K.  
TITLE Molecular cloning and characterization of prostate, an  
androgen-regulated serine protease with prostate-restricted  
expression  
Proc. Natl. Acad. Sci. U.S.A. 96 (6), 3114-3119 (1999)  
JOURNAL MEDLINE 99179024  
REFERENCE 2 (bases 1 to 5900)  
AUTHORS Nelson,P.S., Gan,L., Ferguson,C., Moss,P., Gellinas,R., Hood,L. and  
Wang,K.  
TITLE Direct Submission



Db 2841 CCACAGCTCTACACATCGGGCTGGGCTGCAAGATCTTGAGCGCCAGACAGCAGG 2900  
Qy 171 GAGCCAGATGGTGGAGGCCAGCCTCTCCCTAGGGCCAGAGTACACAGACCTTGT 230  
Db 2901 GAGCCAGATGGTGGAGGCCAGCCTCTCCCTAGGGCCAGAGTACACAGACCTTGT 2960  
Qy 231 CCCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAGTGTGACACCATCG 290  
Db 2961 CCCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAGTGTGACACCATCG 3020  
Qy 291 GAGCATCAGCATTTGCTTCCGATGCCCTACCGCGGGAACTTGTGCTTCTGCTG 350  
Db 3021 GAGCATCAGCATTTGCTTCCGATGCCCTACCGCGGGAACTTGTGCTTCTGCTG 3080  
Qy 351 GGGTCTGCTGGGGAAGGCTGAGCTCAGGGGTGTGTCTGCTTCAAGAGTCTCT 410  
Db 3081 GGGTCTGCTGGGGAAGGCTGAGCTCAGGGGTGTGTCTGCTTCAAGAGTCTCT 3140  
Qy 411 TGCCAGTGGCGGGGCTACCCAGAGCTGTGGTCCAGGAGAAATGCTACCGTGT 470  
Db 3141 TGCCAGTGGCGGGGCTACCCAGAGCTGTGGTCCAGGAGAAATGCTACCGTGT 3200  
Qy 471 CAGTCCGTGAACGTGTGGTGTCTGAGAGGTCTGAGTAAGCTCTATGACCCGCTG 530  
Db 3201 CAGTCCGTGAACGTGTGGTGTCTGAGAGGTCTGAGTAAGCTCTATGACCCGCTG 3260  
Qy 531 TACCAACCCAGATTTCTGCGCCCGGAGGAGGCAAGACCAAGAGCTCTGCAACG 588  
Db 3261 TACCAACCCAGATTTCTGCGCCCGGAGGAGGCAAGACCAAGAGCTCTGCAACG 3318

RESULT 4  
AF148532  
LOCUS AF148532 4385 bp DNA PRI 09-JUN-1999  
DEFINITION Homo sapiens kallikrein 4 (KLK4) gene, complete cds.  
ACCESSION AF148532  
NID 95020095  
VERSION AF148532.1 GI:5020095  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 4385)  
AUTHORS Stephenson, S.A., Vertly, K., Ashworth, L. and Clements, J.A.  
TITLE Localization of a new prostatic specific antigen-related serine  
protease gene, KLK4, is evidence for an expanded human kallikrein  
(KLK) gene family cluster on chromosome 19q13.3-13.4  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 4385)  
TITLE Stephenson, S.A., Ashworth, L. and Clements, J.A.  
JOURNAL Direct Submission  
AUTHORS Submitted (03-MAY-1999) School of Life Science, Queensland  
JOURNAL University of Technology, George Street, Brisbane, Queensland 4001,  
Australia  
FEATURES  
LOCATION/Qualifiers  
SOURCE 1. 4385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13.3-q13.4"  
join(<551..611,1875..2037,2459..2709,2793..2929,4200,  
2400..>4352)  
/gene="KLK4"  
/product="kallikrein 4"  
/gene="KLK4"  
/gene="KLK4"  
join(551..611,1875..2037,2459..2709,2793..2929,4200..4352)  
/gene="KLK4"  
/function="serine protease"  
/codon\_start=1  
/product="kallikrein 4"

/protein\_id="AA038019.1"  
/db\_xref="PID:95020096"  
/db\_xref="GI:5020095"  
/translation="MATAGNMGWFLCYLILGVAGSLVSGSCSOLINGEDCSPHSQPW  
OALVMENELFCSGVLVHPQWVLSAALCFONSTYITGLSLLENDQEPGQWVASIS  
VRHEVYRPLLANDLMLIKLDESVSESDTIRISISIAOCPTAGNSCLVSGWGLANR  
MPVLOCVNVSVSEVCSKLYDPLDLPMSFCAGGGODDSCDNGSDGPIICNGYLO  
GLVEFGAPGQGVGPVYTNLCFTETMIKTYOAS"

BASE COUNT 1013 a 1184 c 1192 g 996 t  
ORIGIN

Query Match 37.5%; Score 474.8; DB 42; Length 4385;  
Best Local Similarity 99.6%; Pred. No. 9,4e-100;  
Matches 476; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 111 CCAGAACTCTTACACCATCGGGCTGGGCTGCACAGTCTTGAAGGCCAGACAGG 170  
Db 2453 CCACAGCTCTACACCATCGGGCTGGGCTGCACAGTCTTGAAGGCCAGACAGG 2512  
Qy 171 GAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGACCCAGAGTACACACCTTGT 230  
Db 2513 GAGCCAGATGGTGGAGGCCAGCCTCTCCGTACGGACCCAGAGTACACACCTTGT 2572  
Qy 231 CCCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAGTGTGACACCATCG 290  
Db 2573 CCCTAACGACCTCATGCTCATCAAGTTGGACGAATCCGTCCGAGTGTGACACCATCG 2632  
Qy 291 GAGCATCAGCATTTGCTTCCGATGCCCTACCGCGGGAACTTGTGCTTCTGCTG 350  
Db 2633 GAGCATCAGCATTTGCTTCCGATGCCCTACCGCGGGAACTTGTGCTTCTGCTG 2692  
Qy 351 GGGTCTGCTGGGGAAGGCTGAGCTCAGGGGTGTGTGCTGCCCTTAAAGAGTCTCT 410  
Db 2693 GGGTCTGCTGGGGAAGGCTGAGCTCAGGGGTGTGTGCTGCCCTTAAAGAGTCTCT 2752  
Qy 411 TGCCAGTGGCGGGGCTGAGCCAGAGCTGTGCGTCCAGGAGATGCTACCGTGTG 470  
Db 2753 TGCCAGTGGCGGGGCTGAGCCAGAGCTGTGCGTCCAGGAGATGCTACCGTGTG 2812  
Qy 471 CAGTCCGTGAACGTGTGGTGTCTGAGAGGTCTGAGTAAGCTCTATGACCCGCTG 530  
Db 2813 CAGTCCGTGAACGTGTGGTGTCTGAGAGGTCTGAGTAAGCTCTATGACCCGCTG 2872  
Qy 531 TACCAACCCAGATTTCTGCGCCCGGAGGAGGCAAGACCAAGAGATCTGCAACG 588  
Db 2873 TACCAACCCAGATTTCTGCGCCCGGAGGAGGCAAGACCAAGAGATCTGCAACG 2930

RESULT 5  
SSU76256  
LOCUS SSU76256 1140 bp mRNA MAM 03-JAN-1998  
DEFINITION Sus scrofa enamel matrix serine proteinase 1 precursor, mRNA,  
complete cds.  
ACCESSION U76256  
NID 92737920  
VERSION U76256.1 GI:2737920  
KEYWORDS  
SOURCE pig.  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
REFERENCE 1 (bases 1 to 1140)  
AUTHORS Simmer, J.P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J.,  
Margolis, H.C., Shimizu, M., Hu, C.-C. and Bartlett, J.D.  
TITLE Purification, Characterization and Cloning of Enamel Matrix Serine  
Proteinase 1  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 1140)  
TITLE Simmer, J.P., Fukae, M., Tanabe, T., Yamakoshi, Y., Uchida, T., Xue, J.,  
Margolis, H.C., Shimizu, M., Hu, C.-C. and Bartlett, J.D.  
JOURNAL Direct Submission  
AUTHORS Submitted (24-OCT-1996) Pediatric Dentistry, University of Texas

Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78284-7888, USA  
Location/Qualifiers

FEATURES  
source 1..1140  
/organism="Sus scrofa"  
/db\_xref="taxon:9823"

## sig\_peptide

69..140  
/note="the signal peptide is cleaved after Ala24 and the preprotein is secreted into the developing enamel matrix"

## CDS

69..833  
/note="EMSP1"  
/codon\_start=1  
/product="enamel matrix serine proteinase 1 precursor"  
/protein\_id="AAB94638.1"  
/db\_xref="GI:2737921"  
/db\_xref="PIR:G2737921"  
/translation="MTAGNPMGWFLLGLITGSLATNGHITNGEDCNPHSOPM QALFLEDFECGVLVHPQWLSAHCFTNSTTGLHNEPEDEPSQHMNASLS IOHPENPSMANDLMLIKLESVSLSDVRNLSVSCPTGSDSLVSGMRLASGR LPQVLQCVNISVAEEVCARYPVHPMFCAGGQDQKDSCHSDSGPLICNSLSQ GLVFGQACGGQNPVGYTNLCKFTDWIQTIIQAS"  
159..830  
/note="the preprotein is cleaved following His30 activating the serine proteinase; the active protein has an apparent molecular weight of 34 kDa and a derived molecular weight of 24 kDa"  
/product="enamel matrix serine proteinase 1"  
order(177..179,549..551)  
/note="disulfide bond"  
order(234..236,282..284)  
/note="disulfide bond"  
join(279..281,414..416,687..689)  
/note="encodes catalytic triad"

## mat\_peptide

159..830  
/note="the preprotein is cleaved following His30 activating the serine proteinase; the active protein has an apparent molecular weight of 34 kDa and a derived molecular weight of 24 kDa"  
/product="enamel matrix serine proteinase 1"  
order(177..179,549..551)  
/note="disulfide bond"  
order(234..236,282..284)  
/note="disulfide bond"  
join(279..281,414..416,687..689)  
/note="encodes catalytic triad"  
405..407  
/note="encodes potential glycosylation location; yields a blank cycle during protein sequencing"  
order(408..410,789..791)  
/note="disulfide bond"  
order(510..512,705..707)  
/note="disulfide bond"  
order(675..677,750..752)  
/note="disulfide bond"  
misc\_count 245 a 366 c 295 g 234 t  
BASE COUNT 245 a 366 c 295 g 234 t  
ORIGIN

Query Match 35.7% Score 452. DB 3; Length 1140;  
Best Local Similarity 74.1% Pred. No. 1.8e-94;  
Matches 640; Conservative 0; Mismatches 140; Indels 84; Gaps 2;

QY 2 GCAGCCCGCACTGGCAGCCCTGGCAGCGGCGACGTGCAATGAAAGAAATGTCGTCT 61  
DB 178 GCAACCCCGCACTGGCAGCCCTGGCAGCGGCGACGTGCAATGAAAGAAATGTCGTCT 237  
QY 62 GGGGCGTCTGTCGATCCGACAGTGGGTGTCAGCCGACACTGTTTCAGAACTCCT 121  
DB 238 GGGGCGTCTGTCGATCCGACAGTGGGTGTCAGCCGACACTGTTTCAGAACTCCT 297  
QY 122 ACACCATCGGCTGGGCTTCACAGTCTTGAAGCCGACAGAGCCAGGAGCCAGATGG 181  
DB 298 ACACCATCGGCTGGGCTTCACAGTCTTGAAGCCGACAGAGCCAGGAGCCAGATGA 357  
QY 182 TGGAGGCGACCTCTCTCGTACGCGACCCAGATACACACCTTCTGCTTAACAGCC 241  
DB 358 TGGAGGCGACCTCTCTCGTACGCGACCCAGATACACACCTTCTGCTTAACAGCC 417  
QY 242 TCATGCTCATCAAGTTGAGAGAAATCCGTCTCGAGTCTGACACCATCCGAGCATCAGCA 301  
DB 418 TCATGCTCATCAAGTTGAGAGAAATCCGTCTCGAGTCTGACACCATCCGAGCATCAGCG 477  
QY 302 TTGCTTGGAGTGGCCCTACCGCGGGGAATCTTGGCTGTTTCTGGCTGGGGTCTGCTGG 361  
DB 478 TCGTCTCCAGTGGCCGAGCCCTGGGGATCTTGGCTGTTTCTGGCTGGGGTCTGCTGG 537

QY 362 CGAAGGTGAGTCTACGGGTGTGTGTCTGCCCTCTTCAAGAGAGTCTCTGCCAGTCCG 421  
DB 538 C----- 539  
QY 422 GGGGCGTACCCAGAGCTGTGCTGCCAGGACAGATGCTTACCGTCTGTCAGTGTGA 481  
DB 539 -----CAGTGCGAGACTGCCCAAGTGTCTGCTGAGTGTGA 574  
QY 482 CGTGTGGGTGTGTCTGAGAGAGTCTGCAATAGCTCTATGACCCGCTGACACCCAG 541  
DB 575 CATCTCGGGCTTCTGAGAGAGTCTGCAAGGCGCTACGCGCCGCTGTACACCCAG 634  
QY 542 CATGTTCTGGCGCGGCGAGGCGACAGACAGAGACTCTGCAAGGAGTCTGTGGG 601  
DB 635 CATGTTCTGGCGCGGCGAGGCGACAGACAGAGACTCTGCAAGGAGTCTGTGGG 654  
QY 602 GCCCTGATCTGCACAGGGGTACTTGCAGGGGCTGTGTCTTGGAAAAAGCCCGTGTG 661  
DB 695 CCCCTCATCTGCAATGGGTGCTGACAGGGGCTGTGTCTTGGAAAAAGCCCGTGTG 754  
QY 662 CCAAGTTGGCGTCCAGGTGTCTACACCAACTCTGCAATTCATGATGATAGAA 721  
DB 755 CCAACCCAGAGTCCAGGGGTCTACACCAACTCTGCAATTCATGATGATAGAA 814  
QY 722 AACCGTCCAGGCGAGTTACTGTGGGAGTGGGACCATGAATTCAGCCCAATPACA 781  
DB 815 GACCATTCAGGCGAGTTACTGCCAGACTGGATGATGAAAGCAACCCCATGTA 874  
QY 782 TCGTGGGAGGAATTCAGGAATATCTGTCCAG-CCCTCTCCCTCAGGCGCAGAG 840  
DB 875 CTGTGCTGAAGAAATTCAGGAATCTGAGTTCCTCCAGTCCCTGACCCAGAG 934  
QY 841 TCCAGGCCCGCAGCCCTCTCTCC 864  
DB 935 TCTAGACACCCCGCAGCCCTCTCC 958

RESULT 6  
AF019979 1237 bp mRNA ROD 01-JAN-1999  
LOCUS AF019979  
DEFINITION Mus musculus enamel matrix serine proteinase 1 precursor, mRNA,  
complete cds.  
ACCESSION AF019979  
NID 94090846  
VERSION AF019979.1 GI:4090846  
KEYWORDS house mouse.  
SOURCE Mus musculus.  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1237)  
AUTHORS Stimmer,J.  
TITLE Enamel Matrix Serine Proteinase 1 (EMSP1)  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1237)  
AUTHORS Stimmer,J.  
TITLE Direct Submision  
JOURNAL Submitted (18-AUG-1997) Pediatric Dentistry, University of Texas  
Health Science Center at San Antonio, 7703 Floyd Curl Drive, San  
Antonio, TX 78284-7888, USA  
FEATURES  
source 1..1237  
/organism="Mus musculus"  
/strain="Swiss-Webster"  
/db\_xref="taxon:10090"  
/cell\_type="ameloblast"  
sig\_peptide 52..126  
CDS 52..819  
/function="degradation of the enamel matrix during enamel  
maturation"  
/note="EMSP1"  
/codon\_start=1

```

/product="enamel matrix serine proteinase 1 precursor"
/protein_id="AAC98894.1"
/db_xref="pid:84090847"
/db_xref="gi:4090847"
/translation="MAYTARIPWGNFLCLILEVTGASASVSSRIIOGDCSPHSOP
WQALFSDGFSCSVLVHPWVLSAACHLOESYLVIGLHLNKSQSPGSMLEAHL
SIQHPNDFSPANDLMLIKNESYIESNTIRSIPTAOCPPGDTCLVSGMWGLKNG
KLPSLDCVNLVSASEEERLLDPVYHLHSMCAAGGDDQNRSCNGSDGGPIVCNRSL
OGVSMGCKGCKGQPIPSVYTNLCKFTMIQIITQTN"
misc_feature
127..144
/note="encodes propeptide"
mat_peptide
145..816
/product="matrix serine proteinase 1"
BASE COUNT      289 a      386 c      286 g      276 t
ORIGIN
Query Match      24.5%   Score 310; DB 12; Length 1237;
Best Local Similarity 61.6%   Pred. No. 8,1e-62;
Matches 642; Conservative 0; Mismatches 296; Indels 105; Gaps 5;
QY      2  GCAGCCCCGACATCGCAGCCCTGGCAGCGGCGACTGCTATGGAAACGAATTGTTGCT 61
DB      164  GGTCCCCACACCTGCGAACCCCTGGCAGCGGCGACTGTTCTCAGAAAGCGGTTTCTGCT 223
QY      62  CGGGCGTCTGTGTCATCGCAGTGGGTGCTGTCAGCCGACACATGTTCCAGAACTGCT 121
DB      224  CGGGAATCTTGTGTCATCGCAGTGGGTGCTGTCAGCAGCAGCACTTCTTCAAGAGATCT 283
QY      122  ACACCATCGGGCTGGGCTGACACAGTCTTGAGCCGACCAAGAGCCAGGAGCCAGATCG 181
DB      284  ACATCGTGGGACTGGGCTGTCATTAACCTGAAGAGGCTCCCAAGAGCTGGCAGCGGATGCT 343
QY      182  TGGAGGCCAGCCTCTCTCCGACGGCAGCCAGAGTACAAAGACCCCTTGCTCCCTAACGACC 241
DB      344  TTAGAGCCCACTCTCTCAACACGACCCCACTTCAATCATCTTCTTGGCAAGATC 403
QY      242  TATGCTCATCAAGTGGAGAAATCCGTGCGAGTGTGACAGCATCGGAGCATCAGCA 301
DB      404  TATGCTCATCAAGTGGAGAAATCCGTGCGAGTGTGACAGCATCGGAGCATCAGCA 463
QY      302  TTGCTTCGAGAGTGGCCTTACCGGGGGAAGTCTTGCTTCTGCTGGGCTGCTGCTG 361
DB      464  TGGTACCCCAATGGCCCTGCTGAGATACGCTAGTCTGCTGGTGGGGTCAACTAA 523
QY      362  CGAAGGTAGCTCAGCGGTGTGTCTGCTCTTCAAGAGAGTCTCTGCCAAGTCC 421
DB      524  AGAATGGGAAC----- 536
QY      422  GGGGCTGACCCAGAGCTGTGCTGCCAGGAGAAATGCTTACCGTGTGAGGGTAA 481
DB      536  -----TGCCAGGCTCTCTCAGTGTGTA 560
QY      482  CGTTCGGTGTGTGAGAGAGTGTGAGTATGATGACCCGCTGTACACCCAG 541
DB      561  TCTTCAGTGGGGGTGAGAGAGTGTGCGGGGTGCTGTATGACCCGTGTACCACTCAG 620
QY      542  CATGTTCTCGCCGGGAGGAGGAGCAAGAGACTCTTGCACAGGCTGACTTGGGG 601
DB      621  TATGTTCTCGCTGGAGGAGAGCAAAAGACTCTTGCACAGGCTGACTTGGGG 680
QY      602  GGGCGTATCTGCAAGGGTACTTGGAGGGCTGTGTTCTTGGAAAGCCCGGTGG 661
DB      681  TCCCATAGTGTGCAGAGATCCCTTACAGGGCTCGTGTATGGGACAAAGAAAGTGG 740
QY      662  CCAAGTGGCGTGCAGAGTGTGTACACCAACTCTGCAATTCAGTGTGATAGAA 721
DB      741  GAGAGCTGCATACCAAGTGTGTACACCAATCTCTGCAAGTGTCACTAAGTGTGAGAC 800
QY      722  AACCTGCCAGGCAAGTAACTCTGGGGAGTGGAAACCATGAATGACCCCAATACA 781
DB      801  CATCAITTCAGACAACTACTATG-----ACTGACTACAAAGGCCACATGAA 848

```

```

QY      782  TCTCGGAGAGAAATTCAGGAATATCTGT--CCAGGCCCTCTCTCCCTCAGGCCAGCA 839
DB      849  GCTGCTGAGGGAACCTGAGACTCAGATGTGCTCTTACCGCTTCTCTCAGATCTAAC 908
QY      840  GTTCAGGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 899
DB      909  GGTCAAGGCCCTTGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 963
QY      900  CTCACACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 959
DB      964  CTCACACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
QY      960  CTCCTCCNTCAGACCCAGAGTCCAGACCCCGCCAGCCCTCTCTCTCTCTCTCTCTCT 1019
DB      1021  CAATACCTCTTTCAGATTCAGAGATTCATGCTCAGACCTCTCTCTCTCTCTCTCTCT 1080
QY      1020  TGAGGCCCGCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1042
DB      1081  GGAACCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1103

RESULT 7
AC005782      AC005782      35197 bp      DNA      PRI      05-OCT-1998
LOCUS      Homo sapiens chromosome 19, cosmid R31855, complete sequence.
DEFINITION
AC005782
AC005782
93702286
VERSION
AC005782.1 GI:3702296
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 35197)
Lamerdin,J.E., McCreedy,P.M., Skowronski,E., Adamson,A.W.,
Burkhardt-Schultz,K., Gordon,L., Kyle,A., Ramirez,M., Stillwagen,S.,
Phan,H., Velasco,N., Do,L., Regala,W., Terry,A., Gaines,J.,
Danganan,L., Poundstone,P., Christensen,M., Georgescu,A., Avila,J.,
Liu,S., Attix,C., Andreise,T., Frankheim,M., Amico-Keller,G.,
Coefield,J., Duarte,S., Lucas,S., Bruce,R., Thomas,P., Quan,G.,
Kronmiller,B., Arellano,A., Montomery,M., Ow,D., Nolan,M.,
Trong,S., Kobayashi,A., Olsen,A.S. and Carrano,A.V.
Sequence analysis of an ~700 kb region in 19q13.4 between D19S268
and SYT3
TITLE
Unpublished
JOURNAL
2 (bases 1 to 35197)
REFERENCE
Lamerdin,J.E.
AUTHORS
Direct Submission
TITLE
Submitted (05-OCT-1998) Joint Genome Institute, Lawrence Livermore
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
JOURNAL
FEATURES
Source
1..35197
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="R31855"
/chromosome="19"
/map="19q13.4 between D19S268 and SYT3"
/clone_11b="LD19NC03 R chromosome 19- specific cosmid
library"
/note="Cosmid library LD19NC03 was constructed at LLNL
from flow-sorted chromosomes from human-hamster hybrid
5H2-B which carries chromosome 19 as its only human
chromosome."
148..255
/note="BLASTX similarity to 008509 (311..346): match:
0.52, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EP88 [pI|539883
EP88 protein - mouse >91|309217 (U21671) EP88 [Mus
155..243
/note="Predicted exon, program: gfall2exons.human_1.3,
frame: 0, quality: excellent, score: 98.000"
251..279
repeat_region

```

```
misc-feature
/ftp_family="GC_rich"
338..430
/Note="BLASTX similarity to 008509 (343..373): match:
0.51, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir||S39983
eps8 protein - mouse >g11309217 (L21671) Eps8 [Mus]
338..412
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
517..660
/Note="BLASTX similarity to 008509 (368..415): match:
0.52, score: 3.3e-23; database searched: nr; EPIDERMAL
GROWTH FACTOR RECEPTOR KINASE SUBSTRATE EPS8 >pir||S39983
eps8 protein - mouse >g11309217 (L21671) Eps8 [Mus]
517..665
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000--DDS
similarity to overlapping ESTs:-(611..665) R34158
y183h09.s1 Homo sapiens cDNA clone 136385 3' similar to
SP:S39983 S39983 EPS8 PROTEIN: (1..55); 100%
identity:--(621..665) R80263 y196h06.s1 Homo sapiens cDNA
clone 147131 3' (1..45); 100% identity."
complement(1145..1258)
/ftp_family="FLMW.C"
1441..1582
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 75.000--DDS
similarity to overlapping ESTs:--R34158 y183h09.s1 Homo
sapiens cDNA clone 136385 3' similar to SP:S39983 S39983
EPS8 PROTEIN: (56..194); 94% identity:--R80263 y196h06.s1
Homo sapiens cDNA clone 147131 3' (46..186); 94%
identity:--H13945 y108d11.s1 Homo sapiens cDNA clone
148149 3' similar to SP:S39983 S39983 EPS8 PROTEIN:
(21..161); 94% identity."
complement(1621..1652)
/ftp_family="GC_rich"
1679..1707
/Note="DSS similarity to overlapping ESTs:--R34158
y183h09.s1 Homo sapiens cDNA clone 136385 3' similar to
SP:S39983 S39983 EPS8 PROTEIN: (195..223); 97%
identity:--R80263 y196h06.s1 Homo sapiens cDNA clone
147131 3' (187..215); 97% identity:--H13945 y108d11.s1
Homo sapiens cDNA clone 148149 3' similar to SP:S39983
S39983 EPS8 PROTEIN: (162..190); 93% identity."
complement(1880..1975)
/ftp_family="MSR1"
3160..3240
/ftp_family="MIR"
complement(3241..3527)
/ftp_family="AluSc"
3507..3702
/ftp_family="MIR"
3904..4036
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 90.000--DDS
similarity to overlapping ESTs:-(3904..3988) R80263
y196h06.s1 Homo sapiens cDNA clone 147131 3' (216..300);
96% identity:--(3904..4036) H13945 y108d11.s1 Homo sapiens
cDNA clone 148149 3' similar to SP:S39983 S39983 EPS8
PROTEIN: (191..323); 95% identity."
3989..4036
/Note="DSS similarity to R34158 y183h09.s1 Homo sapiens
cDNA clone 136385 3' similar to SP:S39983 S39983 EPS8
PROTEIN: (224..271); 100% identity."
4124..4255
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 83.000--DDS
similarity to:-(4124..4215) R34158 y183h09.s1 Homo
sapiens cDNA clone 136385 3' similar to SP:S39983 S39983
EPS8 PROTEIN: (272..364); 94% identity:--(4124..4136)
H13945 y108d11.s1 Homo sapiens cDNA clone 148149 3'
similar to SP:S39983 S39983 EPS8 PROTEIN: (324..336); 100%
identity."

misc-feature
4455..4581
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 95.000"
4777..4951
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 93.000"
complement(4817..4878)
/ftp_family="GC_rich"
5366..5498
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 93.000--DDS
similarity to:-(5367..5498) H13944 y108d11.r1 Homo sapiens
cDNA clone 148149 5' similar to SP:S39983 S39983 EPS8
PROTEIN: (368..236); 93% identity."
5591..5677
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 89.000--DDS
similarity to:-(5591..5825) H13944 y108d11.r1 Homo sapiens
cDNA clone 148149 5' similar to SP:S39983 S39983 EPS8
PROTEIN: (235..1); 99% identity."
complement(5619..5988)
/Note="BLASTN similarity to A1126400 (11..380): match:
0.99, score: 5.0e-144; database searched: month.na;
gc55f11.x1 Soares.placenta.8to9weeks.2NBHP8to9W Homo
sapiens cDNA clone IMAGE:1713549 3', mRNA sequence [Homo
sapiens]
complement(5691..5961)
/Note="DSS similarity to R34157 y183h09.r1 Homo sapiens
cDNA clone 136385 5'. Score: 513 Identity: 268/270
(99%)."
5852..5983
/Note="BLASTN similarity to D25793 (1..132): match: 0.98,
score: 2.8e-45; database searched: est. human clone
3 directed M01 cDNA, HMGSO4161, clone cm1029."
complement(6867..6897)
/ftp_family="GC_rich"
7429..8057
/ftp_family="LINE2"
complement(8297..8602)
/ftp_family="AluX"
8879..9294
/Note="DSS similarity to T23821 seq1081 Homo sapiens cDNA
clone hAB3MA-cot51.5-HAP-Ft-23 3'. Score: 597 Identity:
311/319 (97%).--Additional EST matches: A1123316, T33476
and many others". (9621)
complement(9535..9621)
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 92.000"
complement(9677..9763)
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000"
complement(10285..10389)
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000"
complement(10580..10651)
/ftp_family="GC_rich"
complement(10685..10725)
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 85.000"
complement(10810..10931)
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 100.000--DDS
similarity to R35625 y966g06.r1 Homo sapiens cDNA clone
35310 5' (254..318); 91% identity."
complement(11216..11296)
/Note="predicted exon, program: grail2exons_human_1.3,
frame: 0, quality: excellent, score: 100.000--DDS
```











```

repeat_region      /rpt_family="AluJo"
                    1360. .1398
                    /rpt_family="(TGA)n"
repeat_region      complement(1399. .1483)
                    /rpt_family="MER67C"
repeat_region      2634. .2937
                    /rpt_family="AluSx"
repeat_region      3399. .3495
                    /rpt_family="MLT1A1"
repeat_region      complement(3674. .3716)
                    /rpt_family="MIR"
repeat_region      3807. .4109
                    /rpt_family="AluSg"
repeat_region      4129. .4417
                    /rpt_family="AluSg"
repeat_region      4498. .4578
                    /rpt_family="AluJ/FRAM"
repeat_region      4581. .4774
                    /rpt_family="L1MC/D"
repeat_region      complement(5020. .5087)
                    /rpt_family="Alu"
repeat_region      complement(5095. .6281)
                    /rpt_family="L1"
repeat_region      complement(6416. .6557)
                    /rpt_family="L1M4"
repeat_region      complement(6558. .6860)
                    /rpt_family="AluY"
repeat_region      complement(7329. .7457)
                    /rpt_family="MER4_Internal"
repeat_region      7478. .7839
                    /rpt_family="MER50"
repeat_region      complement(7869. .7920)
                    /rpt_family="(TA)n"
repeat_region      7937. .8213
                    /rpt_family="MER4_Internal"
repeat_region      complement(8259. .9153)
                    /rpt_family="L1PA2"
repeat_region      9684. .9975
                    /rpt_family="AluJo"
repeat_region      11171. .11311
                    /rpt_family="MIR"
repeat_region      complement(12081. .12261)
                    /rpt_family="MIR"
repeat_region      12362. .12397
                    /rpt_family="MLT1D"
repeat_region      12399. .12691
                    /rpt_family="AluSg"
repeat_region      12892. .12869
                    /rpt_family="MLT1D"
repeat_region      complement(13008. .13038)
                    /rpt_family="AluSg"
repeat_region      complement(13074. .13218)
                    /rpt_family="MER1A"
repeat_region      complement(13256. .13442)
                    /rpt_family="MER1A"
repeat_region      complement(13622. .13993)
                    /rpt_family="HERVK"
repeat_region      complement(14138. .14301)
                    /rpt_family="AluSg"
repeat_region      complement(14308. .14692)
                    /rpt_family="MER57_Internal"
repeat_region      14855. .15152
                    /rpt_family="AluSp"
repeat_region      complement(15179. .16657)
                    /rpt_family="MER65_Internal"
misc_feature        /note="DDS similarity to overlapping ESTs:-(16426. .16026)
                    W43007 zc424g09.s1 Soares senescent fibroblasts NBHSF Homo
                    sapiens cDNA clone 323296 3' similar to gb:M20882
                    PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN E PRECURSOR
                    (HUMAN); (401. .10); 99% identity.-(16494. .16238) W49630
                    zc42h12.r1 Soares senescent fibroblasts NBHSF Homo sapiens
                    cDNA clone 325031 5' similar to gb:M23575
repeat_region      /rpt_family="L1"
                    complement(19312. .19498)
repeat_region      /rpt_family="MIR"
                    complement(19577. .19678)
repeat_region      19679. .19978
                    /rpt_family="L1"
repeat_region      19981. .20294
                    /rpt_family="AluJb"
repeat_region      20299. .20483
                    /rpt_family="L1"
repeat_region      20467. .22597
                    /rpt_family="L1MB5"
                    complement(23264. .23758)
                    /note="BLASTX similarity to (331. .495); match: 0.89,
                    score: 3.9e-158; database searched: nr; pregnancy-specific
                    beta-1-glycoprotein 11 form r precursor - human"
                    complement(23506. .23688)
                    /note="DDS similarity to W49630 zc42h12.r1 Soares
                    senescent fibroblasts NBHSF Homo sapiens cDNA clone 325031
                    5' similar to gb:M23575 PREGNANCY-SPECIFIC BETA-1
                    GLYCOPROTEIN C PRECURSOR (HUMAN); (1. .183); 99% identity."
                    complement(23507. .23758)
                    /note="BLASTX similarity to (331. .414); match: 0.86,
                    score: 5.7e-104; database searched: nr; pregnancy-specific
                    beta-1-glycoprotein 11 form r precursor - human"
                    complement(23561. .23760)
                    /note="BLASTN similarity to T39371 (75. .274); match: 1,
                    score: 1.7e-95; database searched: est; ya05b01.r1 Homo
                    sapiens cDNA clone 60553 5' similar to SP:A33514 A33514
                    PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN PRECURSOR -"
                    complement(23685. .23760)
                    /note="BLASTN similarity to T39410 (75. .150); match: 1,
                    score: 7.0e-47; database searched: est; ya05f02.r3 Homo
                    sapiens cDNA clone 60603 5' contains Alu repetitive
                    element.."
                    complement(23704. .23760)
                    /note="BLASTN similarity to T49544 (203. .259); match:
                    0.96, score: 1.9e-92; database searched: est; ya76h05.r1
                    Homo sapiens cDNA clone 67641 5' similar to gb:M20881
                    PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN D PRECURSOR"
                    complement(24158. .24362)
                    /note="BLASTN similarity to T49544 (1. .205); match: 0.98,
                    score: 1.9e-92; database searched: est; ya76h05.r1 Homo
                    sapiens cDNA clone 67641 5' similar to gb:M20881
                    PREGNANCY-SPECIFIC BETA-1 GLYCOPROTEIN D PRECURSOR"
                    complement(24158. .24188)
                    /note="BLASTN similarity to T39410 (47. .77); match: 0.96,
                    score: 7.0e-47; database searched: est; ya05f02.r3 Homo
                    sapiens cDNA clone 60603 5' contains Alu repetitive
                    element.."
                    complement(24158. .24188)

```







**This Page Blank (uspto)**



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: September 28, 1999, 12:27:24 ; Search time 289.74 Seconds  
(without alignments)  
1092.334 Million cell updates/sec

Title: US-09-030-606-173  
 Port of: 1265

Sequence: 1 GGCAGCCCGCACTCGCAGCC.....AGAGANGNCAGAAAAAAA 1265

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:1

Pred. No.	Is the number
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

**Pred. NO.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1254	99.1	1265	1	V58645	Prostate tumour sp
2	1254	99.1	1265	1	V61250	CDNA sequence of p
3	1042.4	82.4	1248	1	V58644	Prostate tumour sp
4	1042.4	82.4	1248	1	V61249	CDNA sequence of p
5	1036.2	81.9	1167	1	V58647	Prostate tumour sp
6	1036.2	81.9	1167	1	V61252	CDNA sequence of p
7	871.6	68.9	1386	1	V11855	Homo sapiens Tubb1
8	641	50.7	871	1	V37445	Human prostate-spe
9	474	37.5	1459	1	V58646	Prostate tumour sp
10	474	37.5	1459	1	V61251	CDNA sequence of p
11	361.4	28.6	1119	1	V58648	Prostate tumour sp
12	361.4	28.6	1119	1	V61253	CDNA sequence of p
13	346.8	27.4	402	1	X41114	Human secreted pro
14	184.6	14.6	8174	1	O13332	GDP-Fuc-beta-D-gal
15	184.6	14.6	8174	1	O56908	DNA encoding a gly
16	184.6	14.6	8174	1	T61617	Human alpha(1,2)-F
17	154	12.2	234	1	V58522	Prostate tumour sp
18	154	12.2	234	1	V61168	CDNA sequence of p
19	143	11.3	4067	1	O63192	Human adeno-associ
20	118.4	9.4	1476	1	X16255	Human kerat
21	109.8	8.7	1438	1	T79126	Human serine prote
22	109.8	8.7	1526	1	V07132	Protease M, a nove
23	106.2	8.4	732	1	O53467	DNA encoding zyme
24	96.6	7.6	1089	1	T39783	Human amyloid prec
25	95	7.5	986	1	O81203	Human stratum corne
26	87.4	6.9	90	1	V58593	Prostate tumour sp
27	87.4	6.9	90	1	V61208	CDNA sequence of p
28	87.2	6.9	1146	1	V84589	Human secreted prote
29	86.8	6.9	833	1	V42925	DNA encoding a hum
30	85.8	6.8	693	1	T61861	Human encoding t
31	81	6.4	897	1	T49878	A. contortrix probio
32	76.4	6.2	1003	1	T13316	Porcine trypsinsogen
33	75.4	6.0	741	1	N81673	Korean Viper salmone
34	75.4	6.0	744	1	T03999	Human spleen trypsin
35	75.4	6.0	744	1	T04000	Human pancreatic trypsin
36	75.4	6.0	744	1	T04001	Human pancreatic trypsin
37	75.4	6.0	744	1	V24548	Human pancreatic trypsinogen-like p
38	73.8	5.8	699	1	V16367	Nucleotide sequenc
39	71.4	5.6	683	1	O63794	Bovine trypsin gene
40	70.8	5.6	907	1	N70905	Human kallikrein-I
41	70.4	5.6	925	1	N93196	Monkey recombinant
42	69.8	5.5	701	1	O63785	bovine trypsinogen
43	69.4	5.5	701	1	X23288	Human trypsin

44	69.2	5.5	738	1	N71045
45	68.8	5.4	957	1	Q20501

## ALIGNMENTS

## REST

ID	standard; CDNA; 1265 BP.
V58645	

DT 08-DEC-1998 (first entry)

KW Prostate tumour specific gene; human; p

OS Homo sapiens.

PD 27-AUG-1998.

PR 09-FEB-1998;

PR 01-AUG-1997;

PI Dillon DC, Xu

PT Novel human F

PS Claim 1; Page 1

CC be used in th

CC - able to bind

CC portion of th

CC also be confu

**SQ Sequence 12**

1. The first group of respondents (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The second group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The third group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The fourth group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The fifth group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The sixth group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The seventh group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The eighth group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The ninth group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost. The tenth group (n = 10) was asked to identify the most important factors influencing their decision to use a mobile app. The factors identified were: ease of use, reliability, security, and cost.

Best Local Similarity 100.0%; Pred. No. 3.9e-296;

\_\_\_\_\_

Db	1	GGCACC	CGGCAT	CTCCAG	CCGCTT	CGGACGGG	GGGAC	TGCTG	TATG	GAAGAA	ACGA	ATTG	TTCT	CG	60
Oy		TCGGG	CGTCC	TGTCAT	CTCCG	CAGTGGG	GGTGT	CAG	CCG	CAC	AGTTT	CCAG	ACTC	120	
Db	61	TCGGG	CGTCT	TGTGCAT	CTCCG	CAGTGGG	GGTGT	CAG	CCG	CAC	AGTTT	CCAG	ACTC	120	
Oy	121	TACAC	CTATG	GGGCT	GGGCT	GCACAG	TTCTT	GAG	GGC	CAC	AAG	CCG	AGG	180	
Db	121	TACAC	CTATG	GGGCT	GGGCT	GCACAG	TTCTT	GAG	GGC	CAC	AAG	CCG	AGG	180	
Oy	181	GTGAG	GGCCAG	GCCTT	CCGT	TACGG	CAC	CCAG	AGTAC	ACAG	ACCC	TTGCT	CGCTAAC	240	
Db	181	GTGAG	GGCCAG	GCCTT	CCGT	TACGG	CAC	CCAG	AGTAC	ACAG	ACCC	TTGCT	CGCTAAC	240	
Oy	241	CTCAT	GCTCAT	TAAGTT	TGGAC	GAATCC	GTGT	CCG	AGTCT	GAC	ACCA	CTCCG	GAG	300	
Db	241	CTCAT	GCTCAT	TAAGTT	TGGAC	GAATCC	GTGT	CCG	AGTCT	GAC	ACCA	CTCCG	GAG	300	
Oy	301	ATTG	TGTGG	CAGTGG	CCAT	ACGCG	GGG	GAAC	TTTGG	CTCTG	TTTGG	GGG	GTG	360	
Db	301	ATTG	TGTGG	CAGTGG	CCAT	ACGCG	GGG	GAAC	TTTGG	CTCTG	TTTGG	GGG	GTG	360	
Oy	361	GGCA	ACGGT	GTAG	ACTAC	AGG	TGTGT	GTG	CGCC	CTT	CAG	GAG	GTCT	420	
Db	361	GGCA	ACGGT	GTAG	ACTAC	AGG	TGTGT	GTG	CGCC	CTT	CAG	GAG	GTCT	420	
Oy	421	CGGG	GGCTG	ACCCAG	AGCTT	CTG	CTGCC	AGG	CAG	ATATG	CCAT	ACCG	TGCTG	480	
Db	421	CGGG	GGCTG	ACCCAG	AGCTT	CTG	CTGCC	AGG	CAG	ATATG	CCAT	ACCG	TGCTG	480	

```

Db      |||||
421 CGGGGGCTACCCAGAGCTCTGCTCCAGGCGAATGCCCTACCGTGTGATGCGTGA 480
Oy      |||||
481 ACGTGTGCTGTGTGTGAGAGGTCTGAGTATGACCGCGTGTACCCACCA 540
Db      |||||
481 ACGTGTGCTGTGTGTGAGAGGTCTGAGTATGACCGCGTGTGTACCCACCA 540
Oy      |||||
541 GCATGTTCTGCGCGCGGAGGAGGCAAGACAGAGACTCTGCAACGGTGTGAGG 600
Db      |||||
541 GCATGTTCTGCGCGCGGAGGAGGCAAGACAGAGACTCTGCAACGGTGTGAGG 600
Oy      |||||
601 GCGCCCTGATCTGCAACGGGTACTTCCAGGCGCTGTGTCTTGGAAAAAGCCG 660
Db      |||||
601 GCGCCCTGATCTGCAACGGGTACTTCCAGGCGCTGTGTCTTGGAAAAAGCCG 660
Oy      |||||
661 GCGAAGTTGGCGTGGCGGTCTTACCAACCTCTGCAATTCAGTGTGATGAGA 720
Db      |||||
661 GCGAAGTTGGCGTGGCGGTCTTACCAACCTCTGCAATTCAGTGTGATGAGA 720
Oy      |||||
721 AAACCGTCCAGGCGCAATTAATCTGTGGGACTGGGAAACCATGAAATGACCCCA 780
Db      |||||
721 AAACCGTCCAGGCGCAATTAATCTGTGGGACTGGGAAACCATGAAATGACCCCA 780
Oy      |||||
781 ATCTGTGCGGAAGGAATTCAGAAATATCTGTCCAGCCCTCTCTCCAGGCGCAG 840
Db      |||||
781 ATCTGTGCGGAAGGAATTCAGAAATATCTGTCCAGCCCTCTCTCCAGGCGCAG 840
Oy      |||||
841 TCCAGGCGCCGAGCGCTCTCTCTCAACCAAGGTACAGATCCCAAGCCCTCTCT 900
Db      |||||
841 TCCAGGCGCCGAGCGCTCTCTCTCAACCAAGGTACAGATCCCAAGCCCTCTCT 900
Oy      |||||
901 TCAGACCCAGAGATCCAGACCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Db      |||||
901 TCAGACCCAGAGATCCAGACCCCGCCCTCTCTCTCTCTCTCTCTCTCTCTCT 960
Oy      |||||
961 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Db      |||||
961 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1020
Oy      |||||
1021 GAGGCGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Db      |||||
1021 GAGGCGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1080
Oy      |||||
1081 CAGACCCAGAGGTNNAGTCTCCAGGCGCTCTCTCTCTCTCTCTCTCTCTCT 1140
Db      |||||
1081 CAGACCCAGAGGTNNAGTCTCCAGGCGCTCTCTCTCTCTCTCTCTCTCTCT 1140
Oy      |||||
1141 TGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
Db      |||||
1141 TGAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1200
Oy      |||||
1201 TTTTCAATTTTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
Db      |||||
1201 TTTTCAATTTTGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
Oy      |||||
1261 AAAAA 1265
Db      |||||
1261 AAAAA 1265

```

## RESULT 2

```

V61250
ID V61250 standard: cDNA: 1265 BP.
AC V61250:
DT 06-JAN-1999 (first entry)
DE cDNA sequence of prostate tumour clone P703 splice variant DE2.
KW Prostate; cancer; tumour; vaccine; Immunogen; clone; ss.
OS Homo sapiens.
PN M09837093-A2.
PD 27-AUG-1998.
PF 25-FEB-1998: US-020956.
PR 09-FEB-1998: US-806099.
PR 25-FEB-1997: US-806099.

```

```

PR 01-AUG-1997; US-904804.
PA (CORI-) CORIXA CORP.
PI Dillon DC, Xu J.
DR MPT: 98-609886/51.
PT Polypeptides comprising immunogenic portions of prostate proteins -
PT used in a vaccine for the treatment of prostate cancer
PS Claim 3; Page 105-106; 130pp; English.
CC The present sequence is a new DNA which encodes an immunogenic
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,
CC can be used as a vaccine for the treatment of prostate cancer. The DNA
CC was identified by analysis of a subtracted cDNA library obtained by
CC subtracting a prostate tumour cDNA expression library with a normal
CC tissue cDNA library.
SQ Sequence 1265 BP; 256 A; 432 C; 321 G; 245 T;

Query Match          99.1%; Score 1254; DB 1; Length 1265;
Best Local Similarity 100.0%; Pred. No. 3,9e-296;
Matches 1265; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 GGCAGCCCGCACTGCGACCCCTGGGAGGCGGCACTGTCATGGAAGAAATGTTCTGC 60
Db      1 GGCAGCCCGCACTGCGACCCCTGGGAGGCGGCACTGTCATGGAAGAAATGTTCTGC 60
Oy      61 TCGGGCTGCTGTGTCATCCGAGTGGGTGCTGTGACGCGGCAACTGTTCCAGAACTCC 120
Db      61 TCGGGCTGCTGTGTCATCCGAGTGGGTGCTGTGACGCGGCAACTGTTCCAGAACTCC 120
Oy      121 TACACCATGCGGCTGGCTTCGACAGTCTTGAAGCCGACCAAGAGCCAGGACCATG 180
Db      121 TACACCATGCGGCTGGCTTCGACAGTCTTGAAGCCGACCAAGAGCCAGGACCATG 180
Oy      181 GTGAGGCGCAGCTCTCTCTGAGGAGCCAGGATCAACACACCTTCTGCTTAACAC 240
Db      181 GTGAGGCGCAGCTCTCTCTGAGGAGCCAGGATCAACACACCTTCTGCTTAACAC 240
Oy      241 CTGATGCTCATCAATGTTGAGAGAAATCCGTGTCCAGTGTGACACCATCGGAGATCAGC 300
Db      241 CTGATGCTCATCAATGTTGAGAGAAATCCGTGTCCAGTGTGACACCATCGGAGATCAGC 300
Oy      301 ATTGCTTGCAAGTGGCTTACCGGCGGGAATCTTGTGCTTCTGTGCTGGGCTGCTG 360
Db      301 ATTGCTTGCAAGTGGCTTACCGGCGGGAATCTTGTGCTTCTGTGCTGGGCTGCTG 360
Oy      361 GCGAAGGTGAGCTACAGGAGTGTGTCTGCCCTTGAAGAGGTCTCTGCCACATCG 420
Db      361 GCGAAGGTGAGCTACAGGAGTGTGTCTGCCCTTGAAGAGGTCTCTGCCACATCG 420
Oy      421 CCGGGCTGACCCAGAGCTTGCCTGCCAGGAGAGATGCTTACCGTGTGAGTGTGAG 480
Db      421 CCGGGCTGACCCAGAGCTTGCCTGCCAGGAGAGATGCTTACCGTGTGAGTGTGAG 480
Oy      481 ACGTGTGCTGTGTCTGAGAGAGTCTGCAAGTATGACCGGCTGTACACACCCA 540
Db      481 ACGTGTGCTGTGTGTGAGAGAGTCTGCAAGTATGACCGGCTGTGTACACACCCA 540
Oy      541 GCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGAGACTCTGCAACGGTACTCTGGG 600
Db      541 GCATGTTCTGCGCGCGGAGGAGGCAAGACCAAGAGACTCTGCAACGGTACTCTGGG 600
Oy      601 GCGCCCTGATCTGCAACGGGTAATCTGAGGCGCTTGTCTTGTGGAAGAAACCCCGTGTG 660
Db      601 GCGCCCTGATCTGCAACGGGTAATCTGAGGCGCTTGTCTTGTGGAAGAAACCCCGTGTG 660
Oy      661 GCGAAGTTGGGCTGCGAGGTGTCTACACCAACTCTGCAAAATTCAGTGTGATGAGA 720
Db      661 GCGAAGTTGGGCTGCGAGGTGTCTACACCAACTCTGCAAAATTCAGTGTGATGAGA 720
Oy      721 AAACCGTCCAGGCGCAATTAATCTGTGGGACTGGGAAACCATGAAATGACCCCAAAATAC 780
Db      721 AAACCGTCCAGGCGCAATTAATCTGTGGGACTGGGAAACCATGAAATGACCCCAAAATAC 780
Oy      781 ATCTGTGCGGAAGGAATTCAGAAATATCTGTCCAGCCCTCTCTCTCAAGCCAGGAG 840

```

```

|||||
Db 781 ATCTGGGGAAGGAATTCAGGAATATCTGTCCAGGCCCTCTCTCCAGGCCAGAG 840
Oy 841 TCCAGGGCCCCAGGCCCTCTCTCTCAACCAAGGTACATATCCCAAGCCCTCTCTCC 900
Db 841 TCCAGGGCCCCAGGCCCTCTCTCTCAACCAAGGTACATATCCCAAGCCCTCTCTCC 900
Oy 901 TCAGACCCAGAGATTCAGACACCCCAAGCCCTCTCTCTAGACCCAGAGATTCAGACCC 960
Db 901 TCAGACCCAGAGATTCAGACACCCCAAGCCCTCTCTCTAGACCCAGAGATTCAGACCC 960
Oy 961 TCTCTCMTAGACCCAGAGATTCAGACACCCCAAGCCCTCTCTCTAGACCCAGAGATTC 1020
Db 961 TCTCTCMTAGACCCAGAGATTCAGACACCCCAAGCCCTCTCTCTAGACCCAGAGATTC 1020
Oy 1021 GAGGCCCCCAAGCCCTCTCTCTCTAGAGATTCAGACACCCCAAGCCCTCTCTCTCTCC 1080
Db 1021 GAGGCCCCCAAGCCCTCTCTCTCTAGAGATTCAGACACCCCAAGCCCTCTCTCTCTCC 1080
Oy 1081 CAGACCCAGAGATTCAGACACCCCTCTCTCTCTAGACCCAGAGATTCAGACCCAGC 1140
Db 1081 CAGACCCAGAGATTCAGACACCCCTCTCTCTCTAGACCCAGAGATTCAGACCCAGC 1140
Oy 1141 TAGATTTTCTCTGACACAGTGCCTCTCTCTCTGACACCCCAAGCCCTCTCTCTCTCT 1200
Db 1141 TAGATTTTCTCTGACACAGTGCCTCTCTCTCTGACACCCCAAGCCCTCTCTCTCTCT 1200
Oy 1201 TTTTCATTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
Db 1201 TTTTCATTTTTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1260
Oy 1261 AAAAA 1265
Db 1261 AAAAA 1265

```

## RESULT 3

V58644

ID V58644 standard: cDNA: 1248 BP.

AC V58644: 08-DEC-1998 (first entry)

DT 08-DEC-1998 (first entry)

DE Prostate tumour specific gene clone DEL.

KM Prostate tumour specific gene: human; prostate cancer; detection;

KW therapy: ss.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 217..696

FT CDS /\*tag- a

PN MO9837418-A2.

PD 27-AUG-1998.

PF 25-FEB-1998: U03690.

PR 09-FEB-1997: US-904809.

PR 25-FEB-1997: US-806596.

PR 01-AUG-1997: US-904809.

PA (CORI-) CORIXA CORP.

PI Dillon DC, Xu J.

PI WPI: 98-48085/41.

DR P-PSD: W69387.

PT Novel human prostate specific tumour protein and fragments - useful

PS Claim 1: Page 112: 141pp: English.

CC This sequence represents a human prostate tumour specific gene, and can

CC be used in the method of the invention. The method is for detecting

CC prostate cancer comprises contacting a biological sample with an agent

CC able to bind an immunogenic portion of a prostate protein (such as

CC encoded by this sequence). An antibody which binds to an immunogenic

CC portion of the prostate protein, and the method can be used to detect,

CC monitor progression of, or treat prostate cancers. The antibody may

CC also be conjugated to a therapeutic agent for use in therapy of prostate

CC cancers.

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

```

Query Match 82.4%; Score 1042.4; DB 1: Length 1248;
Best Local Similarity 91.1%; Pred. No. 1.1e-244;
Matches 1163; Conservative 0; Mismatches 18; Indels 95; Gaps 2;

Oy 2 GCAGCCCGCACTCGCAGCCCTGCGAGCGGCGACTGCTCATAGAAACGAATTGCTGCT 61
Db 29 GCAGCCCGCACTCGCAGCCCTGCGAGCGGCGACTGCTCATAGAAACGAATTGCTGCT 88
Oy 62 CGGGCTCTGCTGCTATCCGAGTGGGTGCTGTAGCGGCACTGTCTTCAGAA----- 117
Db 89 CGGGCTCTGCTGCTATCCGAGTGGGTGCTGTAGCGGCACTGTCTTCAGAAAGTGAG 148
Oy 117 -----CTCTTACACCATCGGGGCTGGCTGACAGCTTTGAGCGGACCAAGAGCAG 169
Db 149 TGCAGAGCTCTTACACCATCGGGGCTGGCTGACAGCTTTGAGCGGACCAAGAGCAG 208
Oy 170 GGAAGCAGATGTTGAGAGCGCAGCTCTCCGACGCCACCAAGATACAGACACCTTGC 229
Db 209 GGAAGCAGATGTTGAGAGCGCAGCTCTCCGACGCCACCAAGATACAGACACCTTGC 268
Oy 230 TCGCTAAGACCTCATGCTATCAAGTTGACGAATCCGTTCCGATGACACCATCC 289
Db 269 TCGCTAAGACCTCATGCTATCAAGTTGACGAATCCGTTCCGATGACACCATCC 328
Oy 290 GGACATCAGCATGCTTTCGAGTGCCTTACCGCGGGGAACTTTCGCTGCTTCTGCT 349
Db 329 GGACATCAGCATGCTTTCGAGTGCCTTACCGCGGGGAACTTTCGCTGCTTCTGCT 388
Oy 350 GGGGTCTGCTGCGACAGCGGTGAGCTACAGGGTGTGTCTGCTCTCTCAAGAGTCT 409
Db 389 GGGGTCTGCTGCGAAC----- 406
Oy 410 CTGCCCAGTGGCGGGGGCTGACCCAGAGCTCTGCTCCAGCGCAAGATGCTTACCTGCT 469
Db 406 -----GGCAGAACTGCTTACCTGCT 425
Oy 470 GCAGTGGGTGAAAGTGTGCGGTGTGTCTGAGAGAGTCTGCAAGTATGACCCGCT 529
Db 426 GCAGTGGGTGAAAGTGTGCGGTGTGTCTGAGAGAGTCTGCAAGTATGACCCGCT 485
Oy 530 GTACACCCCGCAGCATGTTCTGCGCGCGGAGGCGCAAGACAGACATCTCTCAACGG 589
Db 486 GTACACCCCGCAGCATGTTCTGCGCGCGGAGGCGCAAGACAGACATCTCTCAACGG 545
Oy 590 TGACTTGGGGGGCCCTGATCTGCAACGGGTACTTTCAGAGGCTTGTCTTTCGAAA 649
Db 546 TGACTTGGGGGGCCCTGATCTGCAACGGGTACTTTCAGAGGCTTGTCTTTCGAAA 605
Oy 650 AGCCCGGTGGGCGCAAGTGGCGGCGAGGTGTACACCAACTCTGCAAAATTCAGTGA 709
Db 606 AGCCCGGTGGGCGCAAGTGGCGGCGAGGTGTACACCAACTCTGCAAAATTCAGTGA 665
Oy 710 GTGATAGAGAAACCGTCTCAGCCAGCACTTACCTGCGGAGTGGGAGCCCATGAATGA 769
Db 666 GTGATAGAGAAACCGTCTCAGCCAGCACTTACCTGCGGAGTGGGAGCCCATGAATGA 725
Oy 770 CCCCCAATATCTCTGCGGAGGAATTCAAGAAATATCTGTTCCAGCCCTCTCTCTC 829
Db 726 CCCCCAATATCTCTGCGGAGGAATTCAAGAAATATCTGTTCCAGCCCTCTCTCTC 785
Oy 830 AGGCGCAGAGTTCAGAGCGCCCAAGCCCTCTCTCTCAACCAAGGATACAGATCCAG 889
Db 786 AGGCGCAGAGTTCAGAGCGCCCAAGCCCTCTCTCTCAACCAAGGATACAGATCCAG 845
Oy 890 CCCCTCTCTCTCAGACCAAGAGTTCAGACCCCAAGCCCTCTCTCTCAGACCCAGG 949
Db 846 CCCCTCTCTCTCAGACCAAGAGTTCAGACCCCAAGCCCTCTCTCTCAGACCCAGG 905
Oy 950 AGTCAAGCCCTCTCTCTCAGACCAAGAGTTCAGACCCCAAGCCCTCTCTCTCAG 1009
Db 906 AGTCAAGCCCTCTCTCTCAGACCAAGAGTTCAGACCCCAAGCCCTCTCTCTCAG 965
Oy 1010 ACCCAGGGGTGAGAGCGCCCAAGCCCTCTCTCTCAGAGTTCAGAGTTCAGAGCCCAAC 1069

```

|||||  
DB ACCGAGGGGTCAGGGCCCCCAACCCCTCCTCCTCAGACTCAGAGGTCCAAAGCCCCCAAC 1025  
QY 966  
1070 CCTCTTCCCCCAGAGAGGTNNAGTCCAGGCCCTCTTCTCTCAGACCCAGNGGT 1129  
DB 1026 CCTCTTCTCCCCAGAGAGGTCCAGGTCCAGGCCCTCCTCCTCAGACCCAGNGGT 1085  
QY 1130 CCAATGCCACCTAGATTTTCCCTGNNACACAGTCCCTCTGTGNNAGTTGACCAACCT 1189  
DB 1086 CCAATGCCACCTAGACTTCCCTGTACACAGTCCCTCTGTGACAGTTGACCAACCT 1145  
QY 1190 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1249  
DB 1146 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1205  
QY 1250 ANGNCGCAAAAAAAA 1265  
DB 1206 AAGNGCAAAAAAAA 1221  
RESULT 4  
V61249  
ID V61249 standard; cDNA: 1248 BP.  
AC V61249;  
DT 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DEL.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN W0937093.42.  
PD 27-AUG-1998.  
PF 25-FEB-1998: 003492.  
PR 09-FEB-1998: US-020956.  
PR 25-FEB-1998: US-806099.  
PR 01-AUG-1997: US-904804.  
PA (COR1-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR P-PSDB: W71871.  
PT Polypeptides comprising immunogenic portions of prostate proteins -  
PS used in a vaccine for the treatment of prostate cancer  
PS Claim 3, Page 104, 130pp; English.  
CC The present sequence is a new DNA which encodes an immunogenic portion  
CC of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.  
SQ Sequence 1248 BP: 288 A; 424 C; 303 G; 228 T;

Query Match 82.4%; Score 1042.4; DB 1; Length 1248;  
Best Local Similarity 91.1%; Pred. No. 1.1e-244;  
Matches 1103; Conservative 0; Mismatches 18; Indels 95; Gaps 2;

QY 2 GCAGCCGCGACTGCGAGCCCTGCGAGGCGGCGACTGTCATGGAAGGATTTGTGCT 61  
DB 29 GCAGCCGCGACTGCGAGCCCTGCGAGGCGGCGACTGTCATGGAAGGATTTGTGCT 88  
QY 62 CGGGGCTCTGCTGTCATCCGAGTGGGTGCTGTCAGCCGACACTGTTTCCAGAA----- 117  
DB 89 CGGGGCTCTGCTGTCATCCGAGTGGGTGCTGTCAGCCGACACTGTTTCCAGAAAGTGAG 148  
QY 117 -----CTCCTACACATCGGGCTGGGCTGACAGTCTTGAGGGCGGACCAAGACCGAG 169  
DB 149 TCGAGAGCTCTTACACCATCGGGCTGGGCTGACAGTCTTGAGGGCGGACCAAGACCGAG 208  
QY 170 GGAGCCAGATGTGTGAGGCGAGCCCTCTCCGTACGAGCACCAGAGATCAACAGACCTTGC 229  
DB 209 GGAGCCAGATGTGTGAGGCGAGCCCTCTCCGTACGAGCACCAGAGATCAACAGACCTTGC 268  
QY 230 TCGCTAAGCACTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACATCC 289  
DB 269 TCGCTAAGCACTCATGCTCATCAAGTTGGAGCAATCCGTGTCGAGTCTGACACATCC 328

QY 290 GGAGATCAGCATGTGCTTCCAGTGCCTTACCGGGGGAACTTGCTGCTTGTGCT 349  
DB 329 GAGACATCAGCATGTGCTTCCAGTGCCTTACCGGGGGAACTTGCTGCTTGTGCT 388  
QY 350 GGGGTCTGCTGCGAAGCGGTGAGTCAAGGGGTGTGTGTGCTTCCCTTCAAGAGGTCT 409  
DB 389 GGGGTCTGCTGCGAAG-----GGCAGAAATGCTTACCGTGTCT 406  
QY 410 CTGCCAGTCTGGGGGGGCTGACCCAGAGCTTGTGCTCCAGGCAAAATGCTTACCGTGTCT 469  
DB 405 -----GGCAGAAATGCTTACCGTGTCT 425  
QY 470 GCAGTGCCTGAACGATGTGCTGCTGCTGAGGAGGTCTGCAATGCTATGACCCGT 529  
DB 426 GCAGTGCCTGAACGATGTGCTGCTGCTGAGGAGGTCTGCAATGCTATGACCCGT 485  
QY 530 GTACCAACCCAGCATGTGCTGCGCGGCGAGGCGCAAGCAAGAGACTCTGCAACGG 589  
DB 486 GTACCAACCCAGCATGTGCTGCGCGGCGAGGCGCAAGCAAGAGACTCTGCAACGG 545  
QY 590 TGACTCTGGGGGGGCGCCCTGATCTGCAAGGGTACTTGCAGGGCTTGTGTCTTGGAAA 649  
DB 546 TGACTCTGGGGGGGCGCCCTGATCTGCAAGGGTACTTGCAGGGCTTGTGTCTTGGAAA 605  
QY 650 AGCCCGGTGGCCAAAGTTGGGCTGCGAGGTGTCAACCAACCTCTGCAATTCACGA 709  
DB 606 AGCCCGGTGGCCAAAGTTGGGCTGCGAGGTGTCAACCAACCTCTGCAATTCACGA 665  
QY 710 GTGATAGAGAAACCGTCCAGGCGAGTTAACTGTGGGAGTGGGAACCATGAATTA 769  
DB 666 GTGATAGAGAAACCGTCCAGGCGAGTTAACTGTGGGAGTGGGAACCATGAATTA 725  
QY 770 CCCCCAAATACATCTCTGCGGAAGGAATTCAGGAATATCTGTTCCAGCCCTCTCTC 829  
DB 726 CCCCCAAATACATCTCTGCGGAAGGAATTCAGGAATATCTGTTCCAGCCCTCTCTC 785  
QY 830 AGGCCAGGAGTCAAGGCGCCCGCCCTCTCTCAACCAAGGATAGAGATCCCGAG 889  
DB 786 AGGCCAGGAGTCAAGGCGCCCGCCCTCTCTCAACCAAGGATAGAGATCCCGAG 845  
QY 890 CCCCCTCTCTCTCAAGCCAGAGTCCAGAGCCCGCCAGCCCTCTCTCTCTCTCTCTCT 949  
DB 846 CCCCCT 905  
QY 950 AGTCAAGCCCT 1009  
DB 906 AGTCAAGCCCT 965  
QY 1010 ACCCAGGGGTTGAGGCGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1069  
DB 966 ACCCAGGGGTTGAGGCGCCCAACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1025  
QY 1070 CCTCTGTTCCCAAGCAGAGAGTNNAGTCCAGCCCTCTCTCTCTCTCTCTCTCTCTCT 1129  
DB 1026 CCMTATTTCCCAAGCAGAGAGTCCAGGTCCAGCCCTCTCTCTCTCTCTCTCTCTCTCT 1085  
QY 1130 CCAATGCCACCTAGATTTTCCCTGNNACACAGTCCCTCTGTGNNAGTTGACCAACCT 1189  
DB 1086 CCAATGCCACCTAGACTTCCCTGTACACAGTCCCTCTGTGNNAGTTGACCAACCT 1145  
QY 1190 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1249  
DB 1146 TACCACTGGTTTTCATTTTNGTCCCTTCCCTAGATCCAGAAATAAAGTTTAAGAG 1205  
QY 1250 ANGNCGCAAAAAAAA 1265  
DB 1206 AAGNGCAAAAAAAA 1221  
RESULT 5  
V58647  
ID V58647 standard; cDNA: 1167 BP.

AC V58647; 08-DEC-1998 (first entry)  
DE Prostate tumour specific gene clone DE13.  
KW Prostate tumour specific gene; human; prostate cancer; detection;  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT CDS 28..645  
FT /tag= a  
PN MO9837418-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; 003690.  
PR 09-FEB-1998; US-904809.  
PR 25-FEB-1997; US-806596.  
PR 01-AUG-1997; US-904809.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI: 98-480805/41.  
DR P-PSDB; W69388.  
PT Novel human prostate specific tumour protein and fragments - useful for detecting and treating prostate cancers  
PS Claim 1: Page 113; 141pp: English.  
CC This sequence represents a human prostate tumour specific gene, and can be used in the method of the invention. The method is for detecting prostate cancer comprising contacting a biological sample with an agent able to bind an immunogenic portion of a prostate protein (such as encoded by this sequence). An antibody which binds to an immunogenic portion of the prostate protein, and the method can be used to detect, monitor progression of, or treat prostate cancers. The antibody may also be conjugated to a therapeutic agent for use in therapy of prostate cancers.  
SQ Sequence 1167 BP: 242 A: 400 C: 287 G: 222 T:

Query Match 81.9%; Score 1036.2; DB 1: Length 1167;  
Best Local Similarity 91.0%; Pred. No. 3.5e-243;  
Matches 1137; Conservative 0; Mismatches 29; Indels 83; Gaps 1;

QY 14 CGCAGCCCTGGCAGGCGGCACTGGTATGGAAGAAAGATTGTTGCTGGGGGCTCTGG 73  
DB 2 CGCAGCCCTGGCAGGCGGCACTGGTATGGAAGAAAGATTGTTGCTGGGGGCTCTGG 61  
QY 74 TGCATCGCAGTGGGTCTGTACGCGCAGCACTGTTCCAGAACTCCTACACCATGGGGC 133  
DB 62 TGCATCGCAGTGGGTCTGTACGCGCAGCACTGTTCCAGAACTCCTACACCATGGGGC 121  
QY 134 TGGGCGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCAGAGTGGAGGCCAGCC 193  
DB 122 TGGGCGCTGCACAGTCTTGAGGCGGACCAAGAGCCAGGAGCAGAGTGGAGGCCAGCC 181  
QY 194 TCTCCGTACGGCAGCCAGAGTACACAGACCCCTGCTGCTAACGACCTATGCTATCA 253  
DB 182 TCTCCGTACGGCAGCCAGAGTACACAGACCCCTGCTGCTAACGACCTATGCTATCA 241  
QY 254 AGTTGGAGCATCGTGTCCGAGTGTGACACATCCGAGCATCAGATGCTCCGAGT 313  
DB 242 AGTTGGAGCATCGTGTCCGAGTGTGACACATCCGAGCATCAGATGCTCCGAGT 301  
QY 314 GCCCTACCGCGGGGAACCTTGTGCTCTGCTGGGCTGCTGCTGGCAACCGTGGAGC 373  
DB 302 GCCCTACCGCGGGGAACCTTGTGCTCTGCTGGGCTGCTGCTGGCAAC----- 355  
QY 374 TCACGGGTGTGTCTGCTCTTCAAGAGAGTCTCTGCCAGTCCGGGGGCTGACCC 433  
DB 355 ----- 355  
QY 434 AGAGCTGTGGTCCAGCAGAAATGCTACCGTGTGAGTGCCTGAAGCTTCGTTGCT 493  
DB 355 -----GGCAGAAATGCTACCGTGTGAGTGCCTGAAGCTTCGTTGCT 398  
QY 494 GTCTGAGAGTCTGAGTAAAGTCTATGACCCGCTGTACACCCAGCATGTTCTGCGC 553  
DB 399 GTCTGAGAGTCTGAGTAAAGTCTATGACCCGCTGTACACCCAGCATGTTCTGCGC 458

QY 554 CGGCGAGGGCAGACAGAGAGTCTCTGCAACGGTGTGGGGGGCCCTGATCTG 613  
DB 459 CGGCGAGGGCAGACAGAGAGTCTCTGCAACGGTGTGGGGGGCCCTGATCTG 518  
QY 614 CAACGGGTACTTGCAGGGCCCTTGTCTTTCGGAAGAGCCCGCTGCGCAATGGGCGT 673  
DB 519 CAACGGGTACTTGCAGGGCCCTTGTCTTTCGGAAGAGCCCGCTGCGCAATGGGCGT 578  
QY 674 GCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATGAGAAAAACGTCAGGC 733  
DB 579 GCCAGGTGTCTACACCAACCTCTGCAAAATTCAGTGTGATGAGAAAAACGTCAGGC 638  
QY 734 CAGTTAACTGTGGGGAGTGGGAACCCATGAAATTCAGCCCAATATCTCTCTGGGAAG 793  
DB 639 CAGTTAACTGTGGGGAGTGGGAACCCATGAAATTCAGCCCAATATCTCTCTGGGAAG 658  
QY 794 AATTCAGGAATATCTGTCCAGCCCTCTCTCCAGGCGCCAGAGTCCAGGCCCCAG 853  
DB 699 AATTCAGGAATATCTGTCCAGCCCTCTCTCCAGGCGCCAGAGTCCAGGCCCCAG 758  
QY 854 CCCCTCTCTCTCAACCAAGAGGTACAGATCCCGACCCCTCTCTCTCAGACCCAGAG 913  
DB 759 CCCCTCTCTCTCAACCAAGAGGTACAGATCCCGACCCCTCTCTCTCAGACCCAGAG 818  
QY 914 TCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGAC 973  
DB 819 TCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGTCCAGCCCTCTCTCTCAGAC 878  
QY 974 CCAGGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGGTTGAGAGCCCAAC 1033  
DB 879 GCAGGAGTCCAGACCCCGCCAGCCCTCTCTCTCAGACCCAGAGGTTGAGAGCCCAAC 938  
QY 1034 CTTCTCTCTCTCAGAGTCCAGAGTCCAGAGCCCGCCAGCCCTCTCTCTCAGAGT 1093  
DB 939 CTTCTCTCTCTCAGAGTCCAGAGTCCAGAGCCCGCCAGCCCTCTCTCTCAGAGT 988  
QY 1094 NNAGTCCCGAGCCCTCTCTCTCTCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAG 1153  
DB 999 NNAGTCCCGAGCCCTCTCTCTCTCAGAGTCCAGAGTCCAGAGTCCAGAGTCCAG 1058  
QY 1154 NACAGAGTCCCGCCCTCTCTCTCTCTCAGAGTCCAGAGTCCAGAGTCCAGAGT 1213  
DB 1059 TACACAGTCCCGCCCTCTCTCTCTCTCTCAGAGTCCAGAGTCCAGAGTCCAGAGT 1118  
QY 1214 TCCCTTTCCCTAGATCCAGAAATTAAGTTAAGAGANGCAAAAAA 1262  
DB 1119 TCCCTTTCCCTAGATCCAGAAATTAAGTTAAGAGANGCAAAAAA 1167

RESULT 6  
V61252 ID V61252 standard; cDNA: 1167 BP.  
V61252:  
AC 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DE13.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN MO9837093-A2.  
PD 27-AUG-1998.  
PF 25-FEB-1998; 003492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PR 01-AUG-1997; US-904804.  
PA (CORI-) CORIXA CORP.  
PI Dillon DC, Xu J;  
DR WPI: 98-609886/51.  
DR P-PSDB; W71782.  
PT Polypeptides comprising immunogenic portions of prostate proteins - used in a vaccine for the treatment of prostate cancer  
PS Claim 3: Page 107; 130pp: English.  
CC The present sequence is a new DNA which encodes an immunogenic portion of a prostate tumour protein. The encoded immunogen, or the DNA itself,

CC can be used as a vaccine for the treatment of prostate cancer. The DNA  
CC was identified by analysis of a subtracted cDNA library obtained by  
CC subtracting a prostate tumour cDNA expression library with a normal  
CC tissue cDNA library.

50 Sequence 1167 BP; 242 A; 400 C; 287 G; 222 T;

Query Match	81.9%;	Score 1036.2;	DB 1;	Length 1167;
Best Local Similarity	91.0%;	Pred. No. 3.5e-243;		
Matches 1137;	Conservative	0;	Mismatches 29;	Indels 83; Gaps 1

Oy	14	CGCAGCCCTGGCAGGCGGCACTGGTCTATGTAACCAATATTGTTCTCGGGGCTCTGG	73
Db	2	CGCAGCCCTGGCAGGCGGCACTGGTCTATGTAACCAATATTGTTCTCGGGGCTCTGG	61
Oy	74	TGCATCCGCACTGGGTCTCTCAACCGCACACTGTTTCCAGAACTCTACACATCGGGC	133
Db	62	TGCATTCGGCACTGGGTCTCTCTCAACCGCACACTGTTTCCAGAACTCTACACATCGGGC	121
Oy	134	TGGGCTGCACTAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGGAGGCTACCC	193
Db	122	TGGGCTGCACTAGTCTTGAGGCGGACCAAGAGCCAGGAGCCAGATGGTGGAGGCTACCC	181
Oy	194	TCTCCGTACGGCACCACCGAGGTACACAGACCTTGTCTGCTTAACGACTCATGCTATCA	253
Db	182	TCTCCGTACGGCACCACCGAGGTACACAGACTTGTCTGCTTAACGACTCATGCTATCA	241
Oy	254	AGTTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGACATACGATTTGTTGGCAGT	313
Db	242	AGTTGGAGCAATCCGTGTCCGAGTCTGACACCATCCGAGACATACGATTTGTTGGCAGT	301
Oy	314	GCCCTACCGCGGGGAACTGTTGCTCGTTCGTGCTGGGATGCTGCGGGGAACGGGTGAGC	373
Db	302	GCCCTACCGCGGGGAACTGTTGCTCGTTCGTGCTGGGATGCTGCGGGGAACGGGTGAGC	355
Oy	374	TCACGGGTGTGTGTCTGCCCTCTTCAAGAGGTCCTCTGCCAGTCCGCGGGGCTGACCC	433
Db	355	-----	355
Oy	434	AGACCTCTCGGCTCCCAAGGAGAAATGCCCTACCGTCTGCAAGTGGGTACGTTCCGTGCT	493
Db	355	-----GGCAAAATGCCCTACCGGTCGTCGCACTGCGTGAAGGTTCCGTGCTG	398
Oy	494	GTTCTGAGAGGTCTGCACTAAGCTCTATGACCCGCTGTACCAACCCAGCATTTCTGGGC	553
Db	399	GTTCTGAGAGANTCTGCACTAAGCTCTATGACCCGCTGTACCAACCCAGCATTTCTGGGC	458
Oy	554	CGCGGAGGCGCAGACCAAGAAAGACTCTGCAACGGTGACTTGGGGGGCGCTGATGTG	613
Db	459	CGCGGAGGCGCAGACCAAGAAAGACTCTGCAACGGTGACTTGGGGGGCGCTGATGTG	518
Oy	614	CAAGGGGACTTGGAGGGGCTTGATGTCCTTGGGAAAAAGCCCGTGGGCAAGTTGGGCT	673
Db	519	CAAGGGGACTTGGAGGGGCTTGATGTCCTTGGGAAAAAGCCCGTGGGCAAGTTGGGCT	578
Oy	674	GCCAGGTCTTCAACCAACCTCTGCAAAATTCAGTGAAGTGAAGAAAAACCGTCCAGGC	733
Db	579	GCCAGGTGTCTAACCAACCTCTGCAAAATTCAGTGAAGTGAAGAAAAACCGTCCAGNC	638
Oy	734	CAGTTAACTCTGGGGACTGGGAAACCATGAAATTGACCCCCCAATATCATCTGCGGAAG	793
Db	639	CAGTTAACTCTGGGGACTGGGAAACCATGAAATTGACCCCCCAATATCATCTGCGGAAG	698
Oy	794	AATTCAGGAATATGTGTCCAGGCCCTCTGCTCTAGGCGCAGAGGTCCAGGCCCCAG	853
Db	659	AATTCAGGAATATGTGTGTCCAGGCCCTCTCTCTAGGCGCAGAGGTCCAGGCCCCAG	758
Oy	854	CCCCTCTCTCTCTCAAAACCAAGGGGTACAGATCCCCAGCCCTCTCTCTAGACCCAGAG	913
Db	759	CCCCTCTCTCTCTCAAAACCAAGGGGTACAGATCCCCAGCCCTCTCTCTAGACCCAGAG	818
Oy	914	TCGACAGCCCCCAGCCCTCTCTCTCTAGACCCAGAGATCCAGCCCTCTCTCTCAATCAGAC	973
	-		

Db	819	TCACAGCCCCCAGACCCCTTCCTCCTCAGACCCAGAGATCAGCCCTCTCTCCTCAGAC	878
Qy	974	CCAGAGTCACAGACCCCCCAGCCCCCTCTCTCCTCAGACCCAGGGGTGAGGCCCCCAAC	1033
Db	879	GCAGAGATCCAGACCCCCCAGCCCTCCTCCTCAGACCCAGGGGTGAGGGGCCCAAC	938
Qy	1034	CCTCCTCCTTAGATGATGAGATGCAGAGCCCAAGCCCTCGTGTCCCGACCCAGAGAT	1093
Db	939	CCTCCTCCTTAGATGATGAGATGCAGAGCCCAAGCCCTCGTGTCCCGACCCAGAGAT	998
Qy	1094	NNAGTCCAGCCCCCTTCTCCTCCTCAGACCCAGNGTCCATGCACTAGATTTCCCTG	1153
Db	999	NCAGTCCACCCCCCTCTCTCCTCAGACCCAGGGGTGCATGCCACCTAGATTTCCCTG	1058
Qy	1154	NACACAGTCCCTCTGTGTGNNAGTTGACCAACTTACCACTTGATTTTCATTTTGG	1213
Db	1059	TACACAGTCCCTCTGTGTGNNAGTTGACCAACTTACCACTTGATTTTCATTTTGG	1118
Qy	1214	TCCCTTCCCTAGATTCACAAATTAAGTTTAAGNANNGCAAAAAA	1262
Db	1119	TCCCTTCCCTAGATTCACAAATTAAGTTTAAGNANNGCAAAAAA	1167

RESULT 7  
V11855

DT	11-SEP-1998	(first entry)
DE	Homo sapiens	Tub Interactor (htl-1) gene.
KM	Serine protease;	tub Interactor; treatment: obesity; cachexia;
KM	anorexia nervosa;	diabetes; cell cycle progression; apoptosis;
KM	neurodegenerative disease;	Alzheimer's disease; drug screening;
KM	Parkinson's disease;	Huntington's chorea; detection; diagnosis;
KM	amyotrophic lateral sclerosis;	sphino cerebellar degeneration; ss.
OS	Homo sapiens.	
FH	Key	Location/Qualifiers
FT	CDS	2..701
FT		/*tag= a
FT		/product= htl-1 protein
FT		/note= "putative serine protease"
PN	W09812302-A1.	
PD	26-MAR-1998.	
PF	05-SEP-1997;	U15627.
PR	21-JUL-1997;	US-897340.
PR	17-SEP-1996;	US-715032.
PA	(MILL-) MILLENNIUM PHARM INC.	
PI	Errada PR, Gameno CJ;	
DR	MP1; 98-217246/19.	
DR	P-PSDB; W59129.	
PT	Tub Interactor genes	- used to develop products for the treatment
PT	of obesity, cachexia,	anorexia nervosa or related disorders e.g.
PT	diabetes	
PS	Claim 10; Fig 1;	120pp; English.
CC	The sequence is that of the Tub Interactor gene htl-1 which	
CC	codes for a putative serine protease. TI genes function	
CC	in biochemical pathways involved in weight control and	
CC	related disorders. The products can be used for treating	
CC	weight disorders, e.g. obesity, cachexia or anorexia nervosa,	
CC	or a related disorder such as diabetes. The products can	
CC	also be used to modulate cell cycle progression and apoptosis.	
CC	Which can be used for treating neurodegenerative diseases	
CC	which are characterised by apoptosis, including Alzheimer's	
CC	disease, Parkinson's disease, Huntington's chorea, amyotrophic	
CC	lateral sclerosis or sphino cerebellar degenerations. The	
CC	products can also be used for detection, diagnosis and	
CC	drug screening.	
SO	Sequence 1886 BP;	318 A; 490 C; 321 G; 249 T;

Query Match	68.9%	Score 871.6;	DB 1;	Length 1386;
Best Local Similarity	81.6%;	Pred. No. 3,8e+203;		
Matches 1147;	Conservative	3;	Mismatches 25;	Indels 230;
				Gaps 4

```

OY 8 CCGACTCGAGCCCTGGAGGCGGCACTGATGAAAAAGAAATGTTCTGCTGGGCG 67
DB 15 CCGACTCGAGGCGCTGGAGGCGGCGGCACTGATGAAAAAGAAATGTTCTGCTGGGCG 74
OY 68 TCTGTGTGATCCGCAAGTGGTGTCTGACCGGCACACTGTTTCCAGAA----- 117
DB 75 TCTGTGTGATCCGCAAGTGGTGTCTGACCGGCACACTGTTTCCAGAAAGTAGTGACA 134
OY 117 -CTCTACACATGGGGGTGGGCTGACAGTCTTAGAGGCCAGCAAGGCCAGGAGGCG 175
DB 135 GCTCTACACATGGGGGTGGGCTGACAGTCTTAGAGGCCAGCAAGGCCAGGAGGCG 194
OY 176 AGATGGTGGAGGCGCAAGCTCTCCGTACGCGCACCAGAGTACAAAGACCTTGCTGCTA 235
DB 195 AGATGGTGGAGGCGCAAGCTCTCCGTACGCGCACCAGAGTACAAAGACCTTGCTGCTA 254
OY 236 ACGACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCA 295
DB 235 ACGACCTCATGCTCATCAAGTGGAGCAATCCGTGTCGAGTCTGACACCATCCGAGCA 314
OY 296 TCAGATGCTTCGCAAGTGGAGTCCGCGGGAACCTTGCTGCTGCTGGGCTG 355
DB 315 TCAGATGCTTCGCAAGTGGAGTCCGCGGGAACCTTGCTGCTGCTGGGCTG 374
OY 356 TCGTGGCGAAGCGTGGAGTCCAGGGGTGTGTGCTGCTTCACAGAGAGTCTCTGCGCC 415
DB 375 TCGTGGCGAAG----- 386
OY 416 AGTGGCGGGGCTGAGCCAGAGCTCTGCTCCAGGAGCAATGCCGTACGCTGCAGTG 475
DB 386 -----GGCAATAGCCCTTACCGTGTGAGT 411
OY 476 CGTGAACGTGTGGTGTGTCTGAGAGAGTCTGCAATAGCTCTATGACCCTGTACCA 535
DB 412 CGTGAACGTGTGGTGTGTCTGAGAGAGTCTGCAATAGCTCTATGACCCTGTACCA 471
OY 536 CCCAGCATGTTCTGGGCGGCGGAGGCAAGACAGAGAGTCTCTGCAAGGCTGACTC 595
DB 472 CCCAGCATGTTCTGGGCGGCGGAGGCAAGACAGAGAGTCTCTGCAAGGCTGACTC 531
OY 596 TGGGGGGGCGGCTGATCTGCAAGGCTGCTGAGGCGCTGTGTCTTTGGAAAAAGCC 655
DB 532 TGGGGGGGCGGCTGATCTGCAAGGCTGCTGAGGCGCTGTGTCTTTGGAAAAAGCC 591
OY 656 GTGTGGCCAAAGTGGGCTGCAAGGCTCTACACCAACTCTGCAATCTAGTAGTGAT 715
DB 592 GTGTGGCCAAAGTGGGCTGCAAGGCTCTACACCAACTCTGCAATCTAGTAGTGAT 651
OY 716 AGAGAAAAAGCT-CCAGGCGAGTAACTGAGGAGTGGGAAACCCATGAATTTGACCC 774
DB 652 AGAGAAAAAGCTGAGGCGAGTAACTGAGGAGTGGGAAACCCATGAATTTGACCC 711
OY 775 AATATATCTCTCGGAGAGATTCAGAAATATCTGTTCCAGCCCTCTCTCCCTCAGGC 834
DB 712 AATATATCTCTCGGAGAGATTCAGAAATATCTGTTCCAGCCCTCTCTCCCTCAGGC 771
OY 835 CAGAGAGTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 894
DB 772 CAGAGAGTCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 831
OY 895 CCGTCTCTGAGGCGGAGTCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 954
DB 832 CCGTCTCTGAGGCGGAGTCCAGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 891
OY 955 AGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 988
DB 892 AGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 951
OY 988 ----- 988
DB 952 GGGGTTCAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1011
OY 988 -----GCCCGAGCCCTCTC 1000

```

```

DB 1012 AGACCCAGAGAGTCCAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1071
OY 1001 CTCCCTCAGAGCCAGGAGGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1060
DB 1072 CTCCCTCAGAGCCAGGAGGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1131
OY 1061 GCCCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1120
DB 1133 GCCCCCAAGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1191
OY 1121 CCCAGNGTCCAAATCCAGCTAGATTTTCCCTGACAGAGTCCGCGGCGGCGGCGGCGG 1180
DB 1192 CCCAGNGTCCAAATCCAGCTAGATTTTCCCTGACAGAGTCCGCGGCGGCGGCGGCGG 1251
OY 1181 ACCCAACTTACCACTGTTGTTTTCATTTTNGTCCCTTTCCCTGATCCAGAAATATA 1240
DB 1252 ACCCAACTTACCACTGTTGTTTTCATTTTNGTCCCTTTCCCTGATCCAGAAATATA 1311
OY 1241 GTTAAAGANGGCAAAAAAAA 1265
DB 1312 GTTAAAGRAGSGCCAAAAAAA 1336

RESULT 8
V37495
ID V37495 standard; DNM: 871 BP.
AC V37495;
DE 07-SEP-1998 (first entry)
DE Human prostate-specific kallikrein (HPSK) encoding DNA.
KV Prostate-specific kallikrein; HPSK; prostate carcinoma; human;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 31..777
FT     /tag="a
FT     /transl_except= (pos:367..369, aa:Xaa)
FT     /transl_except= (pos:412..414, aa:Xaa)
FT     /transl_except= (pos:424..426, aa:Xaa)
FT     /product= "HPSK protein"
FT     /note= "Xaa = unknown"
PN MO9820117-A1.
PD 14-MAY-1998.
PE 31-OCT-1997; U20051.
PR 05-NOV-1996; US-744026.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Goli SK;
DR WPI: 98-286933/25.
DR P-PSDB: M60592.
PT New isolated prostate-specific kallikrein - used to develop products
PT for diagnosis and treatment of, e.g. prostate carcinoma or benign
PT hyperplasia
PS Claim 5; Fig 1A-C; 68bp; English.
CC This DNA encodes a human prostate-specific kallikrein (HPSK). A host cell
CC containing an expression vector comprising the HPSK nucleic acid sequence
CC can be used to produce the protein recombinantly. The HPSK products can
CC be used for the diagnosis of conditions or diseases associated with
CC expression of HPSK such as prostate carcinoma and benign prostate
CC hyperplasia. Agonists and antagonists which specifically bind to HPSK and
CC modulate its activity can be used for the preparation of treatment of
CC such conditions or diseases. The products can also be used for detection
CC and drug screening, especially for the detection of prostate-specific
CC kallikrein (PSK).
SQ Sequence 871 BP; 166 A; 260 C; 258 G; 184 T;

Query Match 50.7%; Score 641; DB 1; Length 871;
Best Local Similarity 89.0%; Pred. No. 3.9e-147;
Matches 737; Conservative 0; Mismatches 8; Indels 83; Gaps 1;

OY 7 CCGCACTCGAGCCCTGGAGGCGGCACTGATGAAAAAGAAATGTTCTGCTGGGCG 66
DB 127 CAGCATTCACACCTCTGGAGGCGGCGGCACTGATGAAAAAGAAATGTTCTGCTGGGCG 186

```



Oy	67	GTCCTGGTGCATCCGACAGTGCGTGTACAGCGCACACTGTTTCCAGAACTCCTAAC	129
Db	187	GTCCTGGTGCATCCGCACTGGGTCTGTACACGCCACACTGTITCCGAAGTCTTAACC	246
Oy	127	ATCGGGCTGGGCTGCACAGTCTTAGAGGCCACCAAGAGCCAGGAGCCAGATGTGGAG	186
Db	247	ATCGGGCTGGGCTGCACAGTCTTAGAGGCCACCAAGAGCCAGGAGCCAGATGTGGAG	306
Oy	187	GCCAGCTCTCCGTACGCGCACCCAGAGTACACAAGACCTTGTCTCGTAAACGACTCATG	246
Db	307	GCCAGCTCTCCGTACGCGCACCCAGAGTACACAAGACCTTGTCTCGTAAACGACTCATG	366
Oy	247	CTCATCAAGTTGGAGCAATCCGTGTCCGAGTGTACACACATCCGGAGCATCAGATTGCT	306
Db	367	NTCATCAAGTTGGAGCAATCCGTGTCCGAGTGTACACACATCCGGAGNATCAGATTGNT	426
Oy	307	TCCGAGTCCCTTACCGCGGGGAACTCTTCCCTCTTTCGTGGCTGGGCTGTCTGGCGAAC	366
Db	427	TCCGAGTCCCTTACCGCGGGGAACTTCTTCCCTCTTTCGTGGCTGGGCTGTCTGGCGAAC	486
Oy	367	GGTAGCTCACGGGTGTGTCTCTCCCTCTTCAAGAGAGTCTCTGCCAAGTGGGGGG	426
Db	487	-----	487
Oy	427	CTGACCCAGACCTGTGCGTCCCAGACGAATCCCTACCGTGTGCTGAGGTGAACGT	486
Db	487	-----GGCAAGATGCTTACCCTGTCTGCAATGGCTGTAACGTGT	523
Oy	487	CGGTGTGTCTGTAGAGAGGTGTGCAGTAAGCTCTATGACCCGCTGTACACCCAGCATGT	546
Db	524	CGGTGTGTCTGTAGAGAGGTGTGTGAGTCTATGACCCGCTGTACACCCAGCATGT	583
Oy	547	TCTGGCGCGGGGAGGAGCAACAGAAGAGCTCTCAACGTGTGATCTGTGGGGGGCCCC	606
Db	584	TCTGGCGCGGGGAGGAGCAACAGAAGAGCTCTCAACGTGTGATCTGTGGGGGGCCCC	643
Oy	607	TGATCTGCACAGGGTACTTTCGAGGGCCCTGTGTCTTTTCGAAAAAGCCCGTGTGGCCAAG	666
Db	644	TGATCTGCACAGGGTACTTTCGAGGGCCCTGTGTCTTTTCGAAAAAGCCCGTGTGGCCAAG	703
Oy	704	TTGGCGTCCACGGGTGTCTACACCAACTCTGCAAAATTCAGTGAATGTAGAGAAAACCG	763
Db	727	TTCAGGCCAGTAACTCTGGGGAGCTGGGAACCATGAAATTTGAGCCCAATAATCATCTG	786
Oy	764	TTCAGGCCAGTAACTCTGGGGAGCTGGGAACCATGAAATTTGAGCCCAATAATCATCTG	823
Db	824	CGGAAGGAATTCAGGAATATCTGTTCCAGAGCCCTCTCCTCCTCAGGCC	871
Oy	787	CGGAAGGAATTCAGGAATATCTGTTCCAGAGCCCTCTCCTCCTCAGGCC	834
Db	871	CGGAAGGAATTCAGGAATATCTGTTCCAGAGCCCTCTCCTCCTCAGGCC	928
RESULT	9		
V58646	V58646	standard; cDNA; 1459 BP.	
ID	V58646		
AC	V58646		
DT	08-DEC-1998	(first entry)	
DE	Prostate tumour specific gene clone DB5.		
KW	Prostate tumour specific gene; human; prostate cancer; detection;		
KM	therapy; ss.		
OS	Homo sapiens.		
PN	M09837418-A2.		
PD	27-AUG-1998		
PF	25-FEB-1998; U03690.		
PR	09-FEB-1998; US-904809.		
PR	25-FEB-1997; US-806596.		
PR	01-AUG-1997; US-904809.		
PA	(CORI-) CORIXA CORP.		
PI	WILSON DC. Xu J.		
PT	wpil; 98-480805/41.		
NP	Npel human prostate specific tumour protein and fragments - useful		

PT for detecting and treating prostate cancers

PS Claim 1; Page 114; 141pp; English.

CC This sequence represents a human prostate tumour specific gene, and can  
CC be used in the method of the invention. The method is for detecting  
CC prostate cancer comprises contacting a biological sample with an agent  
CC able to bind an immunogenic portion of a prostate protein (such as  
CC encoded by this sequence). An antibody which binds to an immunogenic  
CC portion of the prostate protein, and the method can be used to detect,  
CC monitor progression of, or treat prostate cancer. The antibody may  
CC also be conjugated to a therapeutic agent for use in therapy of prostate  
CC cancers.

SQ Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 37.5%; Score 474; DB 1; Length 1459;  
Best Local Similarity 98.4%; Pred. No.1.8e-106;  
Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

DY 104 ACTGTTTCAGAACGCCATACCACTCGGGGCTGGGCTCAGACTTGTAGAGCCGACCAAG 163  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 25 AGTGAGTCAGAGGCTCCATCACCATCGGGGCTGGGCTCAGAGTTTGAAGCGGCACAAG 84  
  
QY 164 AGCCAGGAGGAGCCAGATGTTGAGAGCCACCCTCTCCGTATGGGACCCAAAGTACAAGAC 223  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 85 AGCCAGGAGGAGCCAGATGTTGAGAGCCACCCTCTCCGTATGGGACCCAAAGTACAAGAC 144  
  
QY 224 CCTGCTGCTPACGACCTCATGCTCATCAAGTTGGAGCAATCGGTGCCAGTGTGACA 283  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 145 CCTTCTGCTPACGACCTCATGCTCATCAAGTTGGAGCAATCGGTGCCAGTGTGACA 204  
  
QY 284 CCATCCGAGGAGCATGAGATTGTTGCGAGTGCCTTACCGGGGGAACCTTGGCTCGTTT 343  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 205 CCATCCGAGGAGCATGAGATTGTTGCGAGTGCCTTACCGGGGGAACCTTGGCTCGTTT 264  
  
QY 344 CTGGCTGGGGGCTGCTGTGCGAAGCGGTGAGCTCACGGGTGTGTGTGCCCTTTTAAGA 403  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 265 CTGGCTGGGGGCTGCTGTGCGAAGCGGTGAGCTCACGGGTGTGTGTGCCCTTTTAAGA 324  
  
QY 404 GGTCCTGTGCCAGTAGTCGGGGGGGCTGACCCAGAGCTGTGCTGCCAGGACAGAAATGCTTAC 463  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 325 GGTCCTGTGCCAGTAGTCGGGGGGGCTGACCCAGAGCTGTGCGHCCAGGACAGAAATGCTTAC 384  
  
QY 464 CGTGTGAGAGTGCAGAACGTGTGGGTGTCTTATAGAGAGTCTCAGTAACTTATGA 523  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 385 CGTGTGAGAGTGCAGAACGTGTGGGTGTCTTATAGAGAGTCTCAGTAACTTATGA 444  
  
QY 524 CCCGCTTACCAACCCACCATATGTTTGTGGCGGGGAGGAGGCAAGACAGAAAGACTCTCG 583  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 445 CCCGCTTACCAACCCACCATATGTTTGTGGCGGGGAGGAGGCAAGACAGAAAGACTCTCG 504  
  
QY 584 CAACG 588  
| | | | |  
Db 505 CAACG 509  
  
RESULT 10  
V61251  
ID V61251 standard; cdna; 1459 BP.  
AC V61251:  
DT 06-JAN-1999 (first entry)  
DE cDNA sequence of prostate tumour clone P703 splice variant DE6.  
KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
OS Homo sapiens.  
PN MO9837093-A2.  
PD 27-Aug-1998.  
PF 25-FEB-1998; U03492.  
PR 09-FEB-1998; US-020956.  
PR 25-FEB-1997; US-806099.  
PA 01-AUG-1997; US-904804.  
PI (CORI-) CORIXA CORP.  
TI Dillion DC, Xu J.  
TR WPI; 98-609886/51.  
XT Polypeptides comprising immunogenic portions of prostate proteins -



PT used in a vaccine for the treatment of prostate cancer  
 PS Claim 3; Page 106; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 can be used as a vaccine for the treatment of prostate cancer. The DNA  
 was identified by analysis of a subtracted cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SO Sequence 1459 BP; 427 A; 328 C; 406 G; 295 T;

Query Match 37.5%; Score 474; DB 1; Length 1459;  
 Best Local Similarity 98.4%; Pred. No. 1.8e-106;  
 Matches 477; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

```

OY 104 ACGTGTTCAGAACTCCTACACCTGCGGCTGGCCCTGCACAGTCTTGAGGCCACCAAG 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25 AGTGAGTGCAGAGCTCTTACACCTGCGGCTGGCCCTGCACAGTCTTGAGGCCACCAAG 84
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 164 AGCCAGGAGCAGATGTGGAGGCGAGCCTTCCTGACGACCCAGAGTACAAACAGAC 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 85 AGCCAGGAGCAGATGTGGAGGCGAGCCTTCCTGACGACCCAGAGTACAAACAGAC 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 224 CATTGCTCGCTAACGACCTCATCTCATAGTTCAGATTCGATCCGAGTCTGACA 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 145 CATTGCTCGCTAACGACCTCATCTCATAGTTCAGATTCGATCCGAGTCTGACA 204
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 284 CCAATCCGAGATAGATTCCTTCGACAGTCCCTACCGGGGAACTCTTGCTCTT 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 CCAATCCGAGATAGATTCCTTCGACAGTCCCTACCGGGGAACTCTTGCTCTT 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 344 CTGGCTGGGCTCTGCTGGCGAAGGCTGAGTACAGGGGTGTCTGCTCCCTTCAAGGA 403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 265 CTGGCTGGGCTCTGCTGGCGAAGGCTGAGTACAGGGGTGTCTGCTCCCTTCAAGGA 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 404 GGTCTCTGCCAGTCCGCGGGGCTGACCCAGAGCTCTGCTCCAGAGTACAGATGCTTAC 463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 325 GGTCTCTGCCAGTCCGCGGGGCTGACCCAGAGCTCTGCTCCAGAGTACAGATGCTTAC 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 464 CGTCTCTCAGTGCCTGAACGTGTGGTGTCTGAGAGAGTCTGCAGTAACTCTATGA 523
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 385 CGTCTCTCAGTGCCTGAACGTGTGGTGTCTGAGAGAGTCTGCAGTAACTCTATGA 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 524 CCGGCTTACACCCAGCATGTCTGTGGCGGCGGAGGGGAGAGAGAGAGTCTCTG 583
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 445 CCGGCTTACACCCAGCATGTCTGTGGCGGCGGAGGGGAGAGAGAGAGTCTCTG 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 584 CAACG 588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 505 CAACG 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 11  
 V58648  
 ID V58648 standard; cDNA; 1119 BP.  
 AC V58648:  
 DT 08-DEC-1998 (first entry)  
 DE Prostate tumour specific gene clone DE14.  
 KW Prostate tumour specific gene; human; prostate cancer; detection;  
 KM therapy; ss.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT CDS 34..528  
 PN WO9837418-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998: US-03690.  
 PR 09-FEB-1998: US-904809.  
 PR 25-FEB-1997: US-806596.  
 PR 01-AUG-1997: US-904809.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI; 98-480805/41.

DR P-PSDB; W69389.  
 PT Novel human prostate specific tumour protein and fragments - useful  
 PS for detecting and treating prostate cancers  
 PS Claim 1; Page 116-117; 141pp; English.  
 CC This sequence represents a human prostate tumour specific gene, and can  
 be used in the method of the invention. The method is for detecting  
 CC prostate cancer comprising contacting a biological sample with an agent  
 CC able to bind an immunogenic portion of a prostate protein (such as  
 CC encoded by this sequence). An antibody which binds to an immunogenic  
 CC portion of the prostate protein, and the method can be used to detect,  
 CC monitor progression of, or treat prostate cancers. The antibody may  
 CC also be conjugated to a therapeutic agent for use in therapy of prostate  
 CC cancers.  
 SO Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 28.6%; Score 361.4; DB 1; Length 1119;  
 Best Local Similarity 99.7%; Pred. No. 4e-79;  
 Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 8 CGCACTGCGAGCCCTGGCAGGCGGCACTGTCATGGAAGAAAGCAATTGTTCTCGGGCG 67
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 CGCACTGCGAGCCCTGGCAGGCGGCACTGTCATGGAAGAAAGCAATTGTTCTCGGGCG 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 68 TCCGTGTCATCCGAGTGGTGTGTCACGCGCACACTGTTCCAGAACTCTACACA 127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 TCCGTGTCATCCGAGTGGTGTGTCACGCGCACACTGTTCCAGAACTCTACACA 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 128 TCGGGCTGGGCTTCGACAGTCTTGAGGCGGACCAAGAGCCAGGAGAGAGTGTGAGG 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 122 TCGGGCTGGGCTTCGACAGTCTTGAGGCGGACCAAGAGCCAGGAGAGAGTGTGAGG 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 188 CCAAGCTCTCCGTACGCGCACCGAGTACAAAGAGCCCTTGCTGCTAACGACTATGC 247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 182 CCAAGCTCTCCGTACGCGCACCGAGTACAAAGAGCCCTTGCTGCTAACGACTATGC 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 248 TCATCAAGTTGGAGGAATCCGTGTCGAGTGTGACACATCCGAGATAGCATTCCT 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 242 TCATCAAGTTGGAGGAATCCGTGTCGAGTGTGACACATCCGAGATAGCATTCCT 301
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 308 CGCAGTCCCTTACCGCGGGGAACTTGTGCTTTCTGTGGCTGGGGTCTGTGGCGAAG 367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 302 CGCAGTCCCTTACCGCGGGGAACTTGTGCTTTCTGTGGCTGGGGTCTGTGGCGAAG 361
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 368 GTG 370
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 362 ATG 364
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 12  
 V61253  
 ID V61253 standard; cDNA; 1119 BP.  
 AC V61253:  
 DT 06-JAN-1999 (first entry)  
 DE cDNA sequence of prostate tumour clone P703 splice variant DE14.  
 KW Prostate; cancer; tumour; vaccine; immunogen; clone; ss.  
 OS Homo sapiens.  
 PN WO9837093-A2.  
 PD 27-AUG-1998.  
 PF 25-FEB-1998: US-03492.  
 PR 09-FEB-1998: US-020956.  
 PR 25-FEB-1997: US-806099.  
 PR 01-AUG-1997: US-904804.  
 PA (CORI-) CORIXA CORP.  
 PI Dillon DC, Xu J;  
 DR WPI; 98-609886/51.  
 DR P-PSDB; W71873.  
 PT Polypeptides comprising immunogenic portions of prostate proteins -  
 PT used in a vaccine for the treatment of prostate cancer  
 PS Claim 3; Page 108-109; 130pp; English.  
 CC The present sequence is a new DNA which encodes an immunogenic portion  
 of a prostate tumour protein. The encoded immunogen, or the DNA itself,  
 CC can be used as a vaccine for the treatment of prostate cancer. The DNA

CC was identified by analysis of a subcloned cDNA library obtained by  
 CC subtracting a prostate tumour cDNA expression library with a normal  
 CC tissue cDNA library.  
 SO Sequence 1119 BP; 248 A; 305 C; 282 G; 284 T;

Query Match 28.6%; Score 361.4; DB 1; Length 1119;  
 Best Local Similarity 99.7%; Pred. No. 4e-79;  
 Matches 362; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 8 CGACATCGACGCGCTGGACGCGGCACTGTGATGGAACGAATGTTCTGTCGGGCG 67  
 DB 2 CGACATCGACGCGCTGGACGCGGCACTGTGATGGAACGAATGTTCTGTCGGGCG 61  
 QY 68 TCCTGTGATCGCGAGTGGGTCTGTACGCCACACTGTTCCAGAACTCTTACACCA 127  
 DB 62 TCCTGTGATCGCGAGTGGGTCTGTACGCCACACTGTTCCAGAACTCTTACACCA 121  
 QY 128 TCGGCTGGGCTGACAGTCTTGAAGCGCAGCAAGAGCCAGGAGCCAGATGTTGAGG 187  
 DB 122 TCGGCTGGGCTGACAGTCTTGAAGCGCAGCAAGAGCCAGGAGCCAGATGTTGAGG 181  
 QY 188 CGAGCTTCCTGACGCGACCCAGAGTACAAAGACCTTGTCTGCTAACGACCTATGC 247  
 DB 182 CGAGCTTCCTGACGCGACCCAGAGTACAAAGACCTTGTCTGCTAACGACCTATGC 241  
 QY 248 TCATCAAGTTGAGCAATCGGTGCTCCAGTCTGACACATCGGAGCATGACATTGCTT 307  
 DB 242 TCATCAAGTTGAGCAATCGGTGCTCCAGTCTGACACATCGGAGCATGACATTGCTT 301  
 QY 308 CGCAGTGCCTTACCGGGGAACTCTTGCTTCTGGCTGGGCTGCTCTGGCGAAGC 367  
 DB 302 CGCAGTGCCTTACCGGGGAACTCTTGCTTCTGGCTGGGCTGCTCTGGCGAAGC 361  
 QY 368 GTC 370  
 DB 362 ATC 364

## RESULT 13

X41114  
 ID X41114 standard; cDNA: 402 BP.

AC X41114;  
 DE 17-JUN-1999 (first entry)  
 KW Human secreted protein; EST: expressed sequence tag; diagnosis;  
 KW forensic; gene therapy; chromosome mapping; signal peptide;  
 KW upstream regulatory sequence; cytokine activity; cell proliferation;  
 KW differentiation; haematopoiesis regulation; tissue growth regulation;  
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;  
 KW thrombolytic; anti-inflammatory; tumour inhibition; ds.  
 OS Homo sapiens.  
 PN M09006548-A2.  
 PD 11-FEB-1999.  
 PF 31-JUL-1998; IB1222.  
 PR 01-AUG-1997; US-905135.  
 PA (GSEST) GENSEST.  
 PI Duclet A, Dumas Mline Edwards J, Lacroix B;  
 DR WPT: 99-153778/13.  
 DR P-PSDB: Y12281.  
 PT New nucleic acids encoding human secreted proteins - obtained from  
 PT cDNA libraries prepared from e.g. liver, ovary, brain, prostate,  
 PT kidney, lung, umbilical cord, placenta and colon tissue  
 PS Claim 1: Page 205; 824pp; English.  
 CC X41094 to X41347 represent 5 expressed sequence tags (ESTs) for human  
 CC secreted proteins, and encode the proteins given in Y12261 to Y12514,  
 CC respectively. The proteins given represent the signal peptide and an  
 CC N-terminal fragment of a secreted protein. The nucleic acid sequences  
 CC can be used for producing secreted human gene products. They can also  
 CC be used to develop products for diagnosis and therapy. The proteins  
 CC obtained may have cytokine activity, cell proliferation/differentiation  
 CC activity, haematopoiesis regulating activity, tissue growth regulating  
 CC activity, reproductive hormone regulating activity, chemotactic/  
 CC activity, chemokinetic activity, chemotactic/

CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/  
 CC ligand activity, anti-inflammatory activity, tumour inhibition activity  
 CC or other activities. The products can be used in forensic, gene therapy  
 CC and chromosome mapping procedures. The sequences can also be used for  
 CC obtaining corresponding promoter sequences. The nucleic acids encoding  
 CC the signal peptide can be used for directing extracellular secretion of  
 CC a polypeptide or the insertion of a polypeptide into a membrane, or  
 CC importing a polypeptide into a cell.  
 SO Sequence 402 BP; 80 A; 125 C; 121 G; 75 T;

Query Match 27.4%; Score 346.8; DB 1; Length 402;  
 Best Local Similarity 96.4%; Pred. No. 1.1e-75;  
 Matches 370; Conservative 0; Mismatches 2; Indels 12; Gaps 1;

QY 2 GGAGCCCGACCTCGACGCGCTGGACGCGGCACTGTGATGGAACGAATGTTCTGCT 61  
 DB 17 GGAGCCCGACCTCGACGCGCTGGACGCGGCACTGTGATGGAACGAATGTTCTGCT 76  
 QY 62 CGGCGCTCTGATGATCGCGAGTGGGTCTGTACGCCACACTGTTCCAGAA----- 117  
 DB 77 CGGCGCTCTGATGATCGCGAGTGGGTCTGTACGCCACACTGTTCCAGAAAGTGAG 136  
 QY 117 -----CTCTACACCATCGGGCTGGGCTGACAGTCTTGAAGCGCAGCAAGAGCCAG 169  
 DB 137 TCAGAGCTCTTACACCATCGGGCTGGGCTGACAGTCTTGAAGCGCAGCAAGAGCCAG 196  
 QY 170 GGAGCCAGTGTGAGAGCGACCTCTGCTACGCGACCCAGAGTACAAAGACCTTGC 229  
 DB 197 GGAGCCAGTGTGAGAGCGACCTCTGCTACGCGACCCAGAGTACAAAGACCTTGC 256  
 QY 230 TCGCTACAGACCTCATGCTCATCAAGTTGAGAACGCTGTCGAGTGTGACACATCC 289  
 DB 257 TCGCTACAGACCTCATGCTCATCAAGTTGAGAACGCTGTCGAGTGTGACACATCC 316  
 QY 290 GGAGCATCAGATTGCTTGCAGTGCCTTACCGCGGGGAACTCTTGCTGTTGCTGCT 349  
 DB 317 GGAGCATCAGATTGCTTGCAGTGCCTTACCGCGGGGAACTCTTGCTGTTGCTGCT 376  
 QY 350 GGGGTCTGCTGGCGAAGGTGAGC 373  
 DB 377 GGGGTCTGCTGGCGAAGGTGAGC 400

## RESULT 14

Q13332/C  
 ID Q13332 standard; DNA: 8174 BP.

AC Q13332;  
 DE 07-NOV-1991 (first entry)  
 KW GDP-Fuc:beta-D-galactoside alpha(1,2)-fucosyltransferase gene.  
 KW glycosyltransferase.  
 OS Homo sapiens.  
 PN M09112340-A.  
 PD 22-AUG-1991.  
 PF 14-FEB-1991; U00899.  
 PR 14-FEB-1990; US-480133.  
 PR 14-FEB-1990; US-479858.  
 PR 12-DEC-1990; US-627621.  
 PA (UNMI) UNIV OF MICHIGAN.  
 PI Lowe JB;  
 DR WPT: 91-267151/36.  
 DR P-PSDB: R13751.  
 PT Isolation of gene conveying post-translational characteristic -  
 PT e.g. the presence of soluble or membrane bound oligo or  
 PT polysaccharide or glycosyltransferase.  
 PS Disclosure: Fig 3; 155pp; English.  
 CC The DNA encodes a protein sequence capable of functioning as a  
 CC GDP- fuc: [beta-D-Gal alpha(1,2)- fucosyltransferase. The sequence  
 CC coded by nucleotides 4782 - 5780 represents the functional protein.  
 CC The enzyme produced by the DNA sequence can be used in enzymatic

CC	fucosylation of chain-terminating galactose residues on lactose	
CC	mannose or neolactose type beta-D-galactoside to alpha-2-L-fucose	
CC	residues. See also Q13330-Q13333.	
SQ	Sequence	8174 BP; 1628 A; 2229 C; 2332 G; 1995 T;

**SQ Sequence 8174 BP; 1628 A; 2229 C; 2322 G; 1995 T;**

Query Match	14.6%	Score 184.6	DB 1	Length 8174
Best Local Similarity	78.2%	Pred. No. 6.7e-36		
Matches 272: Conservative	0	Mismatches 64	Indels 12	Gaps 4

**Oy**    795 ATTCAGGAATATCTGTTCCACGCCCTCCTCCTCAGGCCCGAGAGTCCAGGCCCCCAGC    854  
         | | | | | | | | | | | | | | | | | | | | | |  
**Db**    1025 ACTTAGGACTCACGGCTCCGGCCCCCTCCTCCTCAGACCAGAGGTCCAAGCCCCCTGC    966

**Dy**           855 CCTCTCCTCCCTCAACCGAAG-GTACGATCCCAGCCCCCTCTTCTCTCAGACCAGGAG 913  
||||| |||| | | ||| | ||| ||||| ||||| ||||| |||||  
**Dc**         965 CCCTCTTCCTCAGACCAGAGAGTGACAGGCCCA GCCCTCTCTTTCTCACACCCACGAG 906

Db 905 TCCAGA-TCCCTAGCCCTACTCCCTCAGACCAGAGTCCACACCAAGCTCCTCTC 847

846 CCTCAGACCCAGGAGCCCAAGTTCCCCAGCCCTCTCTCCCTCAGATCCAGGAGTACAGG-787

Db 787 -CCAGACCCTCTCCCTCAGACCAGAGTCCAGGCCCCACCCCTCTCCCTCAGAC 729

Db 728 CCAGGAGTCCAGAGAGCCCGAGCCCTCCTCCTCAGACACAGAGGCGCTA 681

Q56908/c  
ID Q56908 standard; DNA; 8174 BP  
AC Q56908;

DE DNA encoding a glycosyltransferase.  
KW Glycosyltransferase; fucosyltransferase; GDP-Fuc; In vitro; cell; ss.  
KW surface; oligosaccharide; ss.

PN	WO9402616-A	
EH	Key	Location/Qualifiers
ET	cds	4686. .5783
FT		/*tag= a

PU 03-FEB-1994. U06703.  
 PF 20-JUL-1993; US-914281.  
 PR 20-JUL-1992; US-914281.  
 PA (UNMI ) UNIV MICHIGAN.

DR WP1: 94-048874/06.  
DR P-PSDB: R45936.  
PT DNA fragment encoding a glycosyltransferase - can be used for in

PT blood gp. determinants, to protect against transplant rejection  
PS disclosure; Fig 3; 249pp; English.  
CC The sequence is that encoding human glycosyl transferase. The enzyme

CC loss of enzyme activity. It can also be used in in vitro reactions to  
CC modify cell surface oligosaccharide mols. e.g. blood group determinants  
CC See also Q56905-12.  
SQ Sequence A174 BP. 1628 A. 2228 C. 3332 G. 1006 T.

Query Match	14.68 :	Score 184.6 :	DB 1 :	length 8174 :
Best Local Similarity	78.28 :	Pred. No. 6.7e-36 :		

795 ATTGAGGAATATCTGTTCACAGCCCCCTCTCCCTTCAGGCCAGGAATCCAGGCCCCAGC 854

Search completed: September 28, 1999, 12:27:55  
Job time: 5295 sec

Qy	1086	CCAGAGTNNAGGTC	CCAGCCCTCTTC	CNMCAGACC	CAGNGTCCAA	1133
Db	728	CCAGGAGTCCAGAG	CCCCCAGCCCTC	CTCCCTCAGAC	ACAGAAAGCCCTA	681

Search completed: September 28, 1999, 12:27:55  
Job time: 5295 sec

---